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

Alvin D. Jeffery,1 Sharon Hewner,2 Lisiane Pruinelli,3 Deborah Lekan,4 Mikyoung Lee,5 Grace Gao,6 Laura Holbrook7 and Martha Sylvia8

1Department of Veterans Affairs and Vanderbilt University Department of Biomedical Informatics, Nashville, Tennessee, USA, 2Family, Community and Health Systems Science Department, University at Buffalo School of Nursing, Buffalo, New York, USA, 3School of Nursing, University of Minnesota, Minneapolis, Minnesota, USA, 4School of Nursing, University of North Carolina, Greensboro, North Carolina, USA, 5College of Nursing, Texas Woman’s University, Denton, Texas, USA, 6Department of Nursing, St. Catherine University, St. Paul, Minnesota, USA, 7American Sentinel University, Aurora, California, USA and 8College of Nursing, Medical University of South Carolina, Charleston, South Carolina, USA

Corresponding Author: Alvin D. Jeffery, 2525 West End Ave., Suite 1475, Nashville, TN 37203, USA (alvinjeffery@gmail.com)

Received 3 August 2018; Revised 3 October 2018; Editorial Decision 1 November 2018; Accepted 23 November 2018

ABSTRACT

Objective: We sought to assess the current state of risk prediction and segmentation models (RPSM) that focus on whole populations.

Materials: Academic literature databases (ie MEDLINE, Embase, Cochrane Library, PROSPERO, and CINAHL), environmental scan, and Google search engine.

Methods: We conducted a critical review of the literature focused on RPSMs predicting hospitalizations, emergency department visits, or health care costs.

Results: We identified 35 distinct RPSMs among 37 different journal articles (n = 31), websites (n = 4), and abstracts (n = 2). Most RPSMs (57%) defined their population as health plan enrollees while fewer RPSMs (26%) included an age-defined population (26%) and/or geographic boundary (26%). Most RPSMs (51%) focused on predicting hospital admissions, followed by costs (43%) and emergency department visits (31%), with some models predicting more than one outcome. The most common predictors were age, gender, and diagnostic codes included in 82%, 77%, and 69% of models, respectively.

Discussion: Our critical review of existing RPSMs has identified a lack of comprehensive models that integrate data from multiple sources for application to whole populations. Highly depending on diagnostic codes to define high-risk populations overlooks the functional, social, and behavioral factors that are of great significance to health.

Conclusion: More emphasis on including nonbilling data and providing holistic perspectives of individuals is needed in RPSMs. Nursing-generated data could be beneficial in addressing this gap, as they are structured, frequently generated, and tend to focus on key health status elements like functional status and social/behavioral determinants of health.

Key words: population health, risk assessment, decision support techniques, community health planning
INTRODUCTION
The Affordable Care Act has pushed the United States health system to focus on the value of care delivered to patients while transitioning away from a transaction-based system centered on payment for services provided. Success in a value-based system hinges on health care delivery system redesign with integration across multiple settings and care providers as well as a focus on the health of and healthcare delivery for whole populations. Aggregated health-related data points from many individuals are key to understanding the health of whole populations. Population-level aggregated data are used to create risk-prediction models and other data-based tools to develop population health management strategies which address better health, better healthcare experience, and lower costs (key elements of the Triple Aim,3 which were reiterated in the Quadruple Aim).9

BACKGROUND AND SIGNIFICANCE
Population definition
The focus of this article is on populations at the macrolevel or "whole populations," where entire populations are assessed and segmented for varying and targeted interventions aimed at preserving health. This definition is in contrast to populations as sets of individuals with a certain illness or health care need, or only as "high-risk" for certain adverse outcomes.3 Macropopulations may be defined by a specific geographic community or region; members of a health plan; employees; a provider’s catchment area; or an aggregation of individuals with special needs.4 In the United States, macropopulations are most often defined by an entity, such as a health plan, government, employer, or health system that is responsible for the payment or delivery of healthcare services for the individuals within the population. Macropopulations are assessed at the aggregate level to understand healthcare needs and develop systems-level interventions. Individuals within macropopulations can be segmented according to adverse event risk and the need for healthcare services into categories such as: healthy; acutely ill; chronically ill; maternal and infant health; stable with significant disability; short period of decline near death; intermittent exacerbations and sudden death; and long dwindling course/frailty.4,5

Population health management
Population health encompasses an understanding of macrolevel trends in health status within the context of the performance of the health system(s) in which care is delivered and is governed by the physical, behavioral, social, cultural, economic, governmental, policy circumstances in which people live and work.6,7 Population health refers to the distribution of health outcomes within a population, the health determinants that influence distribution, and the policies and interventions that impact the determinants.8 Population health management is "a dynamic approach to healthcare that consists of a variety of interrelated approaches ultimately seeking to improve healthcare quality and optimize healthcare spending" (p. 35)9 and includes a "set of interventions designed to maintain and improve people’s health across the full-continuum of care from low-risk, healthy individuals to high-risk individuals with one or more chronic condition".10 A population health management program is a health system level strategy designed to meet the healthcare needs of a large population with varying levels of health and wellness.

Population health informatics
Population health informatics is the use of health information technology (including hardware and software) in combination with information management concepts to support healthcare delivery for populations.11 Population health informatics targets total populations with operational goals of outreach and prevention and care integration focused in organizations providing population health management with key stakeholders including providers, payment systems, the government, and communities.7 A key function of population health informatics is the integration and analysis of data elements from many large heterogeneous datasets. The resulting aggregated dataset can be used to carry out population assessment, risk prediction, and segmentation.

Whole population risk prediction modeling
Applying statistical and machine learning techniques to the aggregated dataset can be used to predict the risk of adverse outcomes.12 The result of risk prediction modeling is that each individual in an entire population is assigned a score representing the risk of the adverse outcome. Population segmentation uses population health data in isolation or in conjunction with risk prediction scores to assign individuals to clinically meaningful groupings based on demographic, clinical, socioeconomic, environmental, and other characteristics. There is an overlap between whole population risk prediction and segmentation in that both are used to separate whole populations into clinically meaningful categories for intervention implementation. For our purposes, we treated both whole population risk prediction and segmentation models (RPSM) as a joint concept.

We define whole-population RPSMs as those that use people as the denominator. The people within this denominator vary along a continuum from wellness to illness. While RPSMs that are applied in subpopulations (such as predicting adverse outcomes in certain illnesses or from certain surgical procedures) are important for managing the health of special populations, those models do not consider whole populations. Risk prediction models that predict readmissions are an example of the complexity in risk prediction for a population compared to a single high risk group. When readmission risk prediction models use index admissions as their denominator, they are not taking into consideration whole populations and do not use people as a denominator. However, when readmission risk is calculated within a whole population, this is considered whole population risk modeling.

Whole population RPSMs seek to use multiple, integrated sources of data. In contrast to countries with single-payer health systems, integrated data sources are more difficult to apply to US-based whole populations because of disparate systems and lack of interoperability. This lack of integration necessitates a critical review distinct from RPSMs developed, tested, and validated in other countries where data are more consolidated.

Ideally, whole population RPSM inputs strive to measure health determinants of most interest in population health including person-level demographics, morbidities, genetics, relationships with family and communities, economic indicators, and environmental factors. Factors that predispose, enable, or reinforce behaviors contributing to health risk; individuals’ own perception of health risk; the physical and sociocultural environment in which individuals live and work; the macroeconomy of public policy, media, and economy; and the performance of healthcare systems are all important determinants of health-related patient outcomes.
Health-system stakeholders

Data elements for person-level determinants can originate from multiple sources and include: patient-generated data through electronic devices, responses to questionnaires, purchasing patterns and more; health-system generated data through patient care delivery; community and government generated data through environmental and public health assessments; health system quality indicators and other factors; and industry-generated data through purchasing and other transactions. While including all these data elements within RPSMs would be ideal, many RPSMs primarily focus on the inclusion of claims-based data generated from a provider encounter. Expanding to additional frequently generated and readily available data elements could improve RPSM performance.

Nurses form the backbone of population health management strategies. In this realm, nurses assess population needs beyond those of the individual who appears for illness or injury treatment. At an executive and management level, nurses create, execute, and oversee population health strategies and have responsibility for population-based health outcomes. At the patient-care level nurses provide transitional care management; case management; behavioral counseling and health coaching; care co-ordination; and oversight of outreach worker and community health workers.

OBJECTIVE

RPSMs are widely researched and applied in practice in the United States as a means for assigning individuals in whole populations to clinically meaningful groups and risk-avoiding classifications in order to target and deliver appropriate interventions. Given the wide variability in RPSM focus, development, and use, we sought to assess the current state of RPSMs that are developed and/or applied in whole populations. We focused our review on data integrated from multiple sources, populations to which models were applied, predictor variables, target health outcomes, and methods used for model development. Further, although nurses are leaders in population health informatics and population health management interventions, the authors’ experiences and environmental scans suggested a dearth of information in the published literature and practice environments regarding the use of nurse-generated data elements in RPSMs and whether RPSMs incorporate data elements representing nursing healthcare services and interventions. Therefore, we also included a focus on whether nurse-generated data were included in the model, whether nurses contributed to model development, and the degree to which nursing care could be influenced by model outputs. Filling this gap can inform population health risk models in moving towards the inclusion of nurse-generated data and a more person-centered approach to population health management.

METHODS

We conducted a critical review of the literature through an extensive search of the literature and a critical evaluation of its quality. Grant and Booth describe critical reviews as incorporating analysis and conceptual innovation that moves beyond mere description. While a critical review does not have the degree of systematicity, structure, and objectivity found in a systematic review, it can be thought of as a needs assessment that is informative in describing the current status and potential gaps of a specific field. We searched for RPSMs developed to predict hospitalizations, emergency department visits, or health care costs. We required the RPSMs to identify individual risk scores based on population-level input data. We began by developing a conceptual model (Figure 1) to guide the literature search, and we collaborated with a biomedical librarian to assist with search terms and strategies.

We searched for relevant citations within seven biomedical databases and because many RPSMs are developed through the private industry sector, we augmented our database search with an internet search engine. We searched MEDLINE, Embase, Cochrane Library, PROSPERO, and CINAHL databases. We limited results to publication dates between January 2007 and May 2017 and written in English. We excluded qualitative studies and clinical trials. Our search terms comprised words and concepts such as population health, risk prediction model, socioeconomic factors, social determinants of health, registries, preventive health services, health status indicators, whole population, risk factors, risk assessment, decision support techniques, prognosis, and forecasting. Supplementary Material Appendix A contains a detailed MEDLINE search strategy.

The search engine Google is widely regarded as the most popular search engine accounting for more than 75% of internet searches. In an attempt to discover RPSMs produced in the private industry sector, we used Google’s advance search function to mimic the biomedical database search strategy. The single search required inclusion of each of the words population, health, and risk. The exact phrase cost OR emergency OR hospitalization was used in combination with the inclusion of any of the words model, assessment, tool, and product. The Google-generated search phrase was: allintext: population health prediction model OR assessment OR tool OR product cost OR emergency OR hospitalization. We restricted results to pages written in English, updated within the previous year of any file type, without explicit results, and with usage rights listed as “free to use or share.” We downloaded Google results into a Microsoft Excel spreadsheet using Data Miner, a software application publicly available free of charge from Software Innovation Lab. The total number of results scrapped was N = 1108 at which time the free trial limit had been reached. We further restricted results by eliminating results that included text references to locations outside of the United States, resulting in 915 results that qualified for review.

Using pairs of research team members, we reviewed all journal abstracts and Google search results for the following criteria: performed in the United States; examined a whole population with people as the basis (ie across the spectrum from wellness to illness and not defined by a disease state, treatment, or event); used electronic data for defining the population and variables; validated by peer-reviewed research or third party evaluator; included a risk prediction or segmentation model; and the predicted outcome included cost, inpatient admissions, and/or emergency department visits. If discrepancies could not be resolved, the entire team reviewed and agreed on the final determination during multiple regularly scheduled, synchronous meetings.

We selected 84 journal articles and 70 websites for a full review. Each research team member performed a deep review of 8–10
journal articles using a standard review form. Two reviewers critiqued the final set (n = 5) of website sources. The entire group deliberated on any questionable sources. Figure 2 summarizes the number of sources reviewed at each phase of the study.

Analysis
RPSMs, as opposed to journal articles or websites, served as the unit of analysis. We assessed each model to understand the definition of the population in which it is applicable; the outcomes that are impacted; the degree to which data are integrated from multiple sources; the determinants that are included as predictors in the model; the mathematical and statistical methods used for risk prediction and segmentation; whether nurse-generated data are included in the model; whether nurses contributed to development of the model; and whether nursing care delivery is determined based on model outputs. Pairs of authors each independently rated a subset of the sources with respect to quality and relevance using an internally developed rubric (see Supplementary Material Appendix B). We resolved discrepancies in ratings via deliberation until reaching a consensus.

For models found in more than one source, we combined information into one model-level entry for analysis. In the event of discrepancies between sources, we chose to be as widely representative as possible. For example, if one source cited the use of medication information as a predictor while the other source did not, in the model-level analysis, medication information would be associated with the model because it is possible to include it in the model. Additionally, we retained the largest sample size reported.

RESULTS
We identified 35 distinct models among the 37 different sources with most sources (n = 31) being journal articles followed by websites (n = 4) and abstracts (n = 2). We discovered two models (StayWell and ACG-Minnesota), each discussed in 2 sources, and aggregated information from both sources for model-level analysis resulting in 35 models included in the analysis. Population/group sizes ranged from approximately 500 to 3 million (median = 56 188). Table 1 provides abbreviated source descriptions, and

Supplementary Material Appendix C includes additional model and source details.

Population
Most RPSMs defined their population as Health Plan Enrollees (20/35 = 57%). Nine models (26%) used an age-defined population (eg children or greater than 65 years of age). Seven models (20%) used an entire state to define a population while one model comprised an entire county, and one model comprised an entire city.

Outcomes
Most RPSMs focused on predicting hospital admissions (n = 18, 51%), followed by costs (n = 15, 43%) and emergency department visits (n = 11, 31%), with some models predicting more than one of these outcomes. When comparing the outcome-of-interest between models in the public domain (n = 23) versus those models that were proprietary (n = 11), we found a greater emphasis on hospital admissions among publicly available models and a greater emphasis on costs among the proprietary models (Figure 3).

Predictors
RPSM predictors were divided into clinical, demographic and administrative categories. Problem lists (ie structured documentation of patient problems that might not necessarily generate a diagnostic billing code) and laboratory values comprised the most common clinical predictors and were included in 37% and 34% of the models, respectively. However, medications, assessments conducted by both nurses and non-nurses, other, substance use, vital signs, and weight were included in about 15% of the models (Figure 4). Age (82%), gender (77%), and race/ethnicity (43%) comprised the most common demographic predictors, but education, geographic location, and income were included in about 20% of the models (Figure 5). International Classification of Diseases (ICD) Codes comprised the most common administrative predictors (69%) followed by admission history, eligibility for insurance, place of service, and payment information (Figure 6).
| Source author                  | Population                                                                 | Predicted outcomes                                                                 | Relevance to Nursing | Relevance to Health |
|-------------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------------------------|----------------------|---------------------|
| Takahashi et al18             | Health Plan Enrollees, Age-limited (≥18 years)                              | Hospital Admissions, Emergency Department Visits                                     | High                 | High                |
| Takahashi et al19             | Health Plan Enrollees, Age-Limited (adult, age not specified)               | Hospital Admissions                                                                  | High                 | High                |
| Hunold et al20                | Entire State, Age-Limited (≥65 years)                                       | Emergency Department Visits                                                          | Low                  | High                |
| Weinstein et al21             | Health Plan Enrollees, Age not specified (18–64 years old)                  | Costs                                                                               | Moderate             | Moderate            |
| Xing et al22                  | Health Plan Enrollees, Age Limited (18–64 years old)                        | Hospital Admissions, Emergency Department Visits                                     | High                 | High                |
| Jin et al23                   | Entire State, All payers, All diseases, All age groups                      | Emergency Department Visits, Costs                                                  | High                 | High                |
| Crane et al24                 | Hospital readmission                                                        | Hospital Admissions, occurrence of an inpatient stay in the prediction year         | High                 | High                |
| Simpson et al25               | Entire County (multiple), Age Limited (≥65 years)                           | Hospital Admissions, Emergency Department Visits                                     | High                 | High                |
| McAnen et al26                | Health Plan Enrollees, Entire State                                        | Hospital readmission                                                                 | Low                  | High                |
| Silverstein et al27           | Age-Limited (≥65 years)                                                     | Hospital readmission                                                                 | Low                  | High                |
| Gandy et al28                 | Employees of a large commercial insurance company, Age not specified        | Hospital Admissions, Emergency Department Visits                                     | Moderate             | Moderate            |
| Murphy et al29                | Health Plan Enrollees, Age-Limited (≥18 years)                             | Costs                                                                               | Moderate             | Moderate            |
| Simon et al30                 | Health Plan Enrollees, Age-Limited (Children)                               | Hospital Admissions, Emergency Department Visits                                     | High                 | High                |
| Hong et al31                  | Patients in a practice-based research network, Age-Limited (adult, age not specified) | Hospital Admissions, Emergency Department Visits, Cancer screening, A1c, LDL        | High                 | High                |
| Slocum et al32                | Data from Uniform Data System for Medical Rehabilitation, All payers, All age groups | Readmission to acute care from inpatient rehabilitation facilities                  | High                 | High                |
| Blumenthal et al33            | Health Plan Enrollees (Medicare ACO, Commercial, Medicaid risk contract populations), Age Limited (≥18 years) | Hospital Admissions, Emergency Department Visits                                     | Moderate             | Moderate            |
| Drozda et al34                | Health Plan Enrollees, Age not specified (18–64 years old)                 | Hospital Admissions, Emergency Department Visits                                     | High                 | High                |
| Hao et al35                   | All patients within the Maine Health Information Exchange system, All payers, All age groups | Hospital Admissions                                                                 | Moderate             | High                |
| Bowen et al36                 | Employees, Age-Limited (18–64 years old)                                   | Costs                                                                               | Moderate             | Moderate            |
| Goetzl et al37                | Health Plan Enrollees, Age-Limited (18–64 years old), employees of seven companies | Costs, modifiable health risk factors: depression, glucose, BP, weight, tobacco, inactivity, stress, cholesterol, nutrition and alcohol | Moderate             | Moderate            |
| Colombi and Craig Wood38      | Health Plan Enrollees, A large Employer (multi-state), Age Limited (35-74 years old) | Hospital Admissions, Emergency Department Visits                                     | High                 | High                |
| Musich et al39                | Health Plan Enrollees, Age Limited (18–63 years old)                        | Hospital Admissions                                                                 | Moderate             | Moderate            |
| Hu et al40                    | Entire State, All payers, All diseases, All age groups                      | Costs                                                                               | Moderate             | High                |
| Hu et al41                    | Entire State, All payers, All diseases, All age groups                      | Emergency Department Visits                                                          | Moderate             | High                |
| Schertzer42a                  | Age-Limited (older adults, age not specified), Philips Lifeline Personal Emergency Response Service Users | Emergency Department Visits, Costs                                                 | High                 | High                |
| Schmeltz et al43              | Entire Country, Healthcare Cost and Utilization Project Nationwide Inpatient Sample | Heat-related illness—Hospital Admissions                                            | Low                  | High                |
| Fortinsky et al44             | Health Plan Enrollees, Age Limited (≥65 years)                             | Hospital Admissions                                                                 | High                 | High                |
| Gao et al45                   | All VA patients treated in 2 years, All payers, All age groups              | Hospital Admissions                                                                 | High                 | High                |
| Yende et al46                 | Age-Limited (≥45 years); Atherosclerosis Risk in Communities and Cardiovascular Health Studies, internal validation, Health, Aging, and Body Composition Study | Hospital Admissions, admission related to pneumonia                                 | High                 | High                |
| Nyce et al47                  | Health Plan Enrollees, Age Limited (18–64 years old)                       | Costs                                                                               | High                 | High                |
| (continued)                   |                                                                             |                                                                                      |                      |                     |
Quality and clinical relevance

When evaluated at the source-level (ie that of the journal article or website), we rated the majority of sources as being of high (n = 25, 68%) or moderate (n = 7, 19%) quality. At the model-level, all but one model was highly (n = 26, 74%) or moderately (n = 8, 23%) relevant to health. Fewer models were considered moderately (n = 14, 40%) or highly (n = 13, 37%) relevant to nursing, and eight (23%) models had low, if any, relevance to nursing. Sources were inconsistent in their reporting of model performance in a validation set.

Quality and clinical relevance

When evaluated at the source-level (ie that of the journal article or website), we rated the majority of sources as being of high (n = 25, 68%) or moderate (n = 7, 19%) quality. At the model-level, all but one model was highly (n = 26, 74%) or moderately (n = 8, 23%) relevant to health. Fewer models were considered moderately (n = 14, 40%) or highly (n = 13, 37%) relevant to nursing, and eight (23%) models had low, if any, relevance to nursing. Sources were inconsistent in their reporting of model performance in a validation set.

Nursing relevance

Out of 37 studies, six sources collected predictors from nurses’ documentation. Given the inconsistencies in journals’ reporting of authors’ credentials and affiliations, we were unable to reliably determine whether nurses contributed to RPSM development. The reviewers considered predictive models from the majority (n = 29) of sources as having the potential to be used to influence nursing care delivery, and we report the degree to which nursing care could be influenced by RPSM outputs in Table 1. For example, functional status was the most significant predictor of the model for acute care readmissions in Slocum et al study32; functional status is often assessed and documented by nurses. In the model of the Relationship Between Modifiable Health Risk Factors and Medical Expenditures, Absenteeism, Short-Term Disability, and Presenteeism by Goetzel et al’s 2009 study,49 all the risk factors (ie high biometric laboratory values, cigarette and alcohol use, and poor emotional health) are often assessed and documented by nurses. The risk factors to measure differences in health plan costs in Hell and colleagues’ 2009 study included health risk assessment data (ie smoke use, physical activity, and BMI), which were often assessed by nurses. The interventions for controlling such risk factors have often been developed and provided by nurses.

DISCUSSION

Our literature review provides a critical evaluation of RPSMs focused on whole populations with outcomes comprising hospitalizations, costs, and emergency department visits. Of the 35 RPSMs we identified, most models defined their populations as health plan enrollees, had moderate-to-high relevance to health, and were described in sources of moderate-to-high quality. The most common data elements serving as predictors included billing codes, laboratory values, problem lists, age, and gender.

Our finding that the majority of models (57%) use administrative billing data from health plans is not surprising because health plans have the longest history of resources to capture and organize administrative claims data for analysis. Additionally, the eligibility criteria required whole-population data with cost and utilization outcomes. Health plans are the most likely healthcare entities to contain these data because they manage the cost and utilization outcomes for large populations in the United States.

One drawback on the reliance on health plan data is that ICD codes are used to define high risk populations, and these often miss functional, social, and behavioral factors that complicate care and that are of great significance to health. It has been estimated that only 20% of outcomes in the realm of length of life and quality of life are attributed to clinical care with health behaviors, social and economic factors, and the physical environment accounting for the remaining 80%.54 While a definitive direct link has not been made between quality and costs,55 there is evidence that quality, cost, and utilization are interrelated along with the experience of healthcare.1 Recognizing that a relationship exists, it is important to include predictors representing the other 80% of determinants of health outcomes.

| Source author | Population | Predicted outcomes | Relevance to Nursing | Relevance to Health |
|---------------|------------|--------------------|---------------------|---------------------|
| Smith et al48 | Health Plan Enrollees, Age Limited (≥ 18 years) | Hospital Admissions, specifically, heart failure hospitalization within the next year | High | High |
| Goetzel et al49 | Health Plan Enrollees, Age Limited (18–64 years old) | Costs | High | High |
| Hill et al50 | Health Plan Enrollees, Age Limited (≥ 18 years) | Costs | High | High |
| Bertsimas et al51 | Health Plan Enrollees, Age not-specified | Costs | Moderate | Moderate |
| Duncan et al52 | Health Plan Enrollees, Age not-specified | Costs | Low | Moderate |
| Inouye et al52 | Age-Limited (≥ 70 years), All payers | Hospital Admissions | Low | High |
| Robinson53 | Health Plan Enrollees (only those with plans having pharmacy benefit), All age groups | Costs | Low | Low |

*aA source found via an internet search without original model details initially available.*
While the majority of models are relevant to nursing, they are often developed without nurses as part of the development team and seldom use data points derived from nursing care. American Association of Colleges of Nursing (AACN) essentials of entry-level public and population health baccalaureate nursing practice include the ability to use appropriate quantitative data in defining problems and intervening to alleviate them. At the research doctorate level, nurses are prepared to conduct research which can be applied to the development of RPSMs, and at the practice doctorate level, nurses are prepared to translate RPSMs into practice. Nurses are well prepared to develop and translate RPSMs into practice, and with their experience and whole-person approach, can offer unique ways to incorporate data sources previously not considered for RPSMs. This study demonstrates ample opportunity for nurses to take leadership in the predictors used in RPSMs to make them more valuable for nursing practice.

With respect to readiness for implementation into practice, RPSMs fall into two major categories: those that are ready to implement as a software purchase (eg the ACG system) and those that have not been translated into readily available software using data commonly available to a healthcare delivery organization. We found that in RPSMs which are readily available for implementation in practice, the majority of predictors for determining utilization, and costs were from medical care delivery. Several studies have expanded on traditional demographic elements and disease markers by examining predictors such as prescription fill rates, laboratory tests, diagnostic and medication information, and frailty indicators (ie falls, walking difficulty, incontinence, weight loss, malnutrition, vision impairment, cognitive impairment, decubitus ulcer, and lack of social support). Increased adoption of EHRs has made it possible to use novel data elements in RPSMs, and some studies suggest EHR data alone may be acceptable in RPSM models. New areas of investigation are needed in text mining and natural language processing techniques for mining EHR free text to capitalize on these data. It is encouraging that while it is currently more difficult to implement some models into practice, these models do include data from sources representing determinants other than those derived solely from medical care.

A recent National Academy of Medicine report defines six clinical and functional segments that require special attention: children with complex needs, nonelderly disabled, people with multiple chronic and major complex chronic conditions, the frail elderly, and those with advanced illness. Our review provides a picture of the current RPSMs, which can inform future directions for predictive models incorporating nursing data.
A significant limitation of the RPSMs was that demographic characteristics were often limited to age and gender. In addition to suggesting the importance of functional limitations in the segmentation of high-need populations, Long and colleagues\cite{long2016} overlay behavioral risk factors and social determinants of health for all six segments. Thus, a clear direction for future predictive models is to include social and behavioral determinants of health. These determinants could be incorporated into future studies as one way to include social factors using health plan data, along with proxy indicators of low income such as Medicaid eligibility.

A few of the studies used information from a statewide health information exchange. This seems a very productive approach as it includes all payers and all healthcare settings. Linking health information exchange with other data sources such as the census would expand the ability to include social determinants of health as well.

Strengths and limitations

Strengths of our review include leveraging both academic literature databases and an internet search strategy during data collection and integrating a focus on both risk prediction models and population segmentation models. Our review is limited by the inability to critically evaluate the internet-based industry models with the same degree of rigor as the peer-reviewed, journal-based academic models, given the proprietary nature of industry-based models and inconsistent reporting standards. Additionally, several false-negative sources could be present; in our efforts to narrow our search strategy, a few sources we expected to be returned were not present in our final collection of sources. Upon review of previous papers from one of the authors in the context of the search strategy, we discovered the abstract text of two expected papers\cite{bodenheimer2014, vuuk2016} did not specifically include the word “model” in the abstract and were therefore not included. Finally, the internet-based search, while helpful in identifying RPSMs used in health care, might not be fully reproducible. The details of Google’s exact method for returning search results is proprietary information, and one should note that search results automatically consider synonyms and order results based on relevance to the overall search strategy. Further, results can be skewed by user profile data including location, gender, and previous searches. To mitigate this potential influence, the search was conducted through a generic computer user profile using a new and neutral Google account that excluded information including location and gender, following the deletion of all search history and cookies.

CONCLUSION

Our critical review of existing RPSMs has identified a lack of comprehensive models that integrate data from multiple sources for application to whole populations. More emphasis is needed to expand RPSMs to include predictors that span nonbilling data in an effort to provide a more realistic and holistic perspective of individual patients’ risk for poor health outcomes. Including nursing-generated data would be a beneficial first step in advancing this work, as these data are frequently generated, available in a structured format, and tend to focus on key health status elements like functional status and social/behavioral determinants of health.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Journal of the American Medical Informatics Association online.

CONTRIBUTORS

All authors contributed to conception/design of the work, data collection, qualitative analysis/interpretation, critical revision of the article, and final approval of the version to be published. AJ conducted the quantitative analyses. SH and MS created the first draft of the manuscript.

FUNDING

This work was supported by the resources and the use of facilities at the Department of Veterans Affairs, Tennessee Valley Healthcare System. Its contents are solely the responsibility of the authors and do not necessarily represent official views of the Department of Veterans Affairs or the United States government.

Conflict of interest statement. None declared.

REFERENCES

1. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff 2008; 27 (3): 759–69.
2. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014; 12 (6): 573.
3. Vuuk SI, Mayer EK, Darzi A. Patient segmentation analysis offers significant benefits for integrated care and support. Health Aff 2016; 35 (5): 769–75.
4. Kharrazi H, Lasser EC, Yasnoff WA, et al. A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop. J Am Med Inform Assoc 2017; 24 (1): 2–12.
5. Lynn J, Straube BM, Bell KM, et al. Using population segmentation to provide better health care for all: the ‘bridges to health’ model. Milbank Q 2007; 85 (2): 185–208.
6. Hewner S, Seo JY, Gothard SE, et al. Aligning population-based care management with chronic disease complexity. Nurs Outlook 2014; 62 (4): 250–8.
7. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999; 89 (9): 1322–7.
8. Kindig DA. Understanding population health terminology. Milbank Q 2007; 85 (1): 139–61.
9. Nash DB, Reisnnyder J, Fabius RJ, et al. Population Health: Creating a Culture of Wellness. 2015. https://books.google.com/books/about/Population_Health.html?id=R4A89gEACAAJ. Accessed July 12, 2018.
10. Felt-Lisk S, Higgins T. Exploring the Promise of Population Health Management Programs to Improve Health. Washington, DC: Mathematica Policy Research: 2011. https://www.mathematica-mpr.com/our-publications-and-findings/publications/exploring-the-promise-of-population-health-management-programs-to-improve-health. Accessed July 12, 2018.
11. Mandil SH. Health informatics by Salah Mandil. 1989; http://apps.who.int/iris/handle/10665/46976. Accessed July 12, 2018.
12. Hewner S, Sullivan SS, Yu G. Reducing emergency room visits and in-hospitalizations by implementing best practice for transitional care using innovative technology and big data. Worldviews Evidence-Based Nurs 2018; 15 (3): 170–7.
13. Radzyninski S. The concept of population health within the nursing profession. J Prof Nurs 2007; 23 (1): 37–46.
14. Jean-Baptiste D, O’Malley A, Shah T. Population Segmentation and Targeting of Health Care Resources: Findings from a Literature Review. Washington, DC: Mathematica Policy Research: 2017. https://www.mathematica-mpr.com/our-publications-and-findings/publications/population-segmentation-and-targeting-of-health-care-resources-findings-from-a-literature-review. Accessed July 12, 2018.
lic health nursing. 2010; 1–45. http://www.aacnnursing.org/Teaching-Resources/Advanced-Practice-Competencies. Accessed July 12, 2018.

58. Chang H-Y, Richards TM, Shermock KM, et al. Evaluating the impact of prescription fill rates on risk stratification model performance. *Med Care* 2017; 55 (12): 1052–60.

59. Lemke KW, Gudzune KA, Kharrazi H, et al. Assessing markers from ambulatory laboratory tests for predicting high-risk patients. *Am J Manag Care* 2018; 24 (6): e190–5.

60. Kharrazi H, Chi W, Chang H-Y, et al. Comparing population-based risk-stratification model performance using demographic, diagnosis and medication data extracted from outpatient electronic health records versus administrative claims. *Med Care* 2017; 55 (8): 789–96.

61. Kan HJ, Kharrazi H, Leff B, et al. Defining and assessing geriatric risk factors and associated health care utilization among older adults using claims and electronic health records. *Med Care* 2018; 56 (3): 233–9.

62. Dzau VJ, McClellan MB, McGinnis JM, et al. Vital directions for health and health care priorities from a National Academy of Medicine initiative. *J Am Med Assoc* 2017; 317 (14): 1461–70.

63. Hewner S, Casucci S, Castner J. The roles of chronic disease complexity, health system integration, and care management in post-discharge health-care utilization in a low-income population. *Res Nurs Health* 2016; 39 (4): 215–28.

64. Hewner S, Wu Y-WB, Castner J. Comparative effectiveness of risk-stratified care management in reducing readmissions in medicaid adults with chronic disease. *J Healthc Qual* 2016; 38 (1): 3–16.