



Life Expectancy Calculators

June 2016

Prepared for:

Department of Veterans Affairs
Veterans Health Administration
Quality Enhancement Research Initiative
Health Services Research & Development Service
Washington, DC 20420

Prepared by:

Evidence-based Synthesis Program (ESP) Center
Minneapolis VA Medical Center
Minneapolis, MN
Tim Wilt, MD, MPH, Director
Nancy Greer, PhD, Program Manager

Investigators:

Principal Investigator:
Thomas Rector, PhD, PharmD

Co-investigators:

Brent Taylor, PhD
Shahnaz Sultan, MD, MHSc
Aasma Shaukat, MD, MPH
Selcuk Adabag, MD, MS
David Nelson, PhD
Timothy Capecchi, MD

Research Associate:

Roderick MacDonald, MS

NOTE: This publication is for internal use of the Department of Veterans Affairs and should not be distributed outside the agency.



VA | Defining
HEALTH CARE | **EXCELLENCE**
in the 21st Century

PREFACE

The VA Evidence-based Synthesis Program (ESP) was established in 2007 to provide timely and accurate syntheses of targeted healthcare topics of particular importance to clinicians, managers, and policymakers as they work to improve the health and healthcare of Veterans. QUERI provides funding for four ESP Centers, and each Center has an active University affiliation. Center Directors are recognized leaders in the field of evidence synthesis with close ties to the AHRQ Evidence-based Practice Centers. The ESP is governed by a Steering Committee comprised of participants from VHA Policy, Program, and Operations Offices, VISN leadership, field-based investigators, and others as designated appropriate by QUERI/HSR&D.

The ESP Centers generate evidence syntheses on important clinical practice topics. These reports help:

- Develop clinical policies informed by evidence;
- Implement effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures; and
- Set the direction for future research to address gaps in clinical knowledge.

The ESP disseminates these reports throughout VA and in the published literature; some evidence syntheses have informed the clinical guidelines of large professional organizations.

The ESP Coordinating Center (ESP CC), located in Portland, Oregon, was created in 2009 to expand the capacity of QUERI/HSR&D and is charged with oversight of national ESP program operations, program development and evaluation, and dissemination efforts. The ESP CC establishes standard operating procedures for the production of evidence synthesis reports; facilitates a national topic nomination, prioritization, and selection process; manages the research portfolio of each Center; facilitates editorial review processes; ensures methodological consistency and quality of products; produces “rapid response evidence briefs” at the request of VHA senior leadership; collaborates with HSR&D Center for Information Dissemination and Education Resources (CIDER) to develop a national dissemination strategy for all ESP products; and interfaces with stakeholders to effectively engage the program.

Comments on this evidence report are welcome and can be sent to Nicole Floyd, ESP CC Program Manager, at Nicole.Floyd@va.gov.

Recommended citation: Rector T, Taylor BC, Sultan S, Shaukat A, Adabag S, Nelson D, Capecchi T, MacDonald R, Greer, N, Wilt TJ. Life Expectancy Calculators, VA ESP Project #09-009; 2016.

This report is based on research conducted by the Evidence-based Synthesis Program (ESP) Center located at the Minneapolis VA Medical Center, Minneapolis, MN, funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Quality Enhancement Research Initiative. The findings and conclusions in this document are those of the author(s) who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (eg, employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in the report.

TABLE OF CONTENTS

Definitions..... 1

Executive Summary 2

 Introduction..... 2

 Methods 2

 Data Sources and Searches 2

 Study Selection 3

 Data Abstraction, Quality Assessment, Synthesis, and Strength of Evidence..... 3

 Results3

 Key Question 1 3

 Key Question 2 4

 Key Question 3 4

 Discussion..... 4

 Key Findings 4

 Applicability 5

 Research Gaps/Future Research 5

 Conclusions..... 5

Introduction..... 6

Methods..... 10

 Topic Development..... 10

 Search Strategy 10

 Study Selection 10

 Data Abstraction 11

 Quality Assessment..... 11

 Data Synthesis..... 11

 Rating the Strength of Evidence 11

 Peer Review 12

Results 13

 Literature Flow 13

 Reviewed Studies..... 14



Study Quality	16
Key Question 1: Between 2011 and 2016, have there been any additional reports of life expectancy calculators that may have sufficient predictive accuracy for use in adult primary care practice?.....	18
Summary of Findings.....	18
Conclusion for Key Question #1.....	21
Rating for the Strength of Evidence for KQ1: High.....	21
Key Question 2: Of the life expectancy calculators being reviewed, have any external validation studies been published between 2011 and 2016? If yes, what population was studied and what was the predictive accuracy therein?	22
Summary of Findings.....	22
Conclusion for Key Question #2.....	22
Rating for the Strength of Evidence for KQ2: Insufficient.....	23
Key Question 3: What is the clinical use of the mortality prediction models (<i>aka</i> life expectancy calculators), and was there improvement in patient survival times, health-related quality of life, provider-patient communication, patient satisfaction and participation in clinical decisions, or lower healthcare utilization and costs?.....	24
Summary of Findings.....	24
Conclusion for Key Question #3.....	24
Rating for the Strength of Evidence for KQ 3: Insufficient.....	24
Summary and Discussion	25
Key Findings.....	25
Limitations	25
Applicability of Findings to the VA Population.....	26
Research Gaps/Future Research	26
Conclusions.....	27
References	28
Tables	
Table 1. Study Population Characteristics	14
Table 2. Description of Prediction Models.....	16
Table 3. Quality of Included Studies	17
Table 4. Predictive Performance of Models under Review	19
Table 5. Validation of Mortality Prediction Models.....	22

Figures

Figure 1. Conceptual Framework 10
Figure 2. Literature Flow Chart 13

Appendix A. Search Strategy..... 30

Appendix B. Peer Review Comments/Author Responses 31

Appendix C. Evidence Tables 35

Table 1. Study Characteristics 35
Table 2. Model Characteristics and Performance 46

DEFINITIONS

Term	Definition
Life expectancy	The estimated (calculated) average number of years a group of people is expected to live. Most individuals in the group will live longer or shorter than the average life expectancy.
Mortality Prediction Model	A statistical model that uses predictor variables to estimate the probability (risk score) an individual will be alive or deceased at a specified future time.
Risk Groups	Groups formed by categorizing individual estimated probabilities of dying.
Survival (or mortality) curve	Graphical plot of the estimated cumulative probability of surviving (or dying) versus time. Cumulative probabilities are often reported as a percentage.
Median survival time	The time when the cumulative probability of survival (or death) reaches 0.50 (50%). May be used as a proxy for life expectancy because 50% of the people in a group are expected to live longer and 50% shorter than the estimated median survival time.
Validation	Testing a prediction model in a new sample of patients that was not used to develop the model. Often validation is done by randomly splitting a sample of patients into one or more subsamples and using one subsample to develop the model and the other subsample to validate the model. However, this approach may be overly optimistic in regards to future predictive performance because the distributions of predictor variables and mortality tend to be similar in randomly split samples.
Calibration	The difference in the predicted number of deaths as compared to the observed number in each risk group. If the differences are small, the model is well-calibrated to the studied sample.
C-statistic	A measure of how well the prediction model's risk scores discriminate individuals who did or did not die within a specified period of time. C-statistics indicate the ability of a prediction model to rank individuals in concordance with their observed survival times. A model with a C-statistic that's not much better than 0.5 will not predict who will live or die much better than flipping a coin. On the other hand, the closer the C-statistic is to 1.0, the more likely it is that the prediction model can be used to make accurate survival predictions with an acceptably low number of prediction errors.
Sensitivity	The proportion of all decedents during a period of time that had risk scores exceeding a threshold being used to predict death.
Specificity	The proportion of survivors during a period of time that had risk scores less than a threshold being used to predict survival.
Positive Predictive Value	Given a proposed risk score threshold for making prognostic predictions, the proportion of patients above the threshold that would be predicted to die within a specified period of time and do.
Negative Predictive Value	Given a proposed risk score threshold for making prognostic predictions, the proportion of patients below the threshold that would be predicted to survive for a specified period of time and do.

EXECUTIVE SUMMARY

INTRODUCTION

Estimation of life expectancy serves an important role in clinical decisions about screening for disease and treatment in primary care practices. Increasingly, clinical guidelines recommend a course of action in context of the patient's life expectancy. While many tools have been developed to estimate life expectancy or the likelihood of surviving for a period of time, there are no well-established decision support tools that are broadly applicable to primary care.

For a life expectancy calculator to gain widespread use in primary care, it needs to have acceptable and validated predictive accuracy at decision-relevant times such as 6 months or 5 years, proven usefulness, and be readily available. In 2012, a systematic review by Yourman et al summarized existing life expectancy calculators that may be useful for older patients in primary care. However, none of the reviewed calculators met our criteria for widespread use. The purpose of this systematic review was to update the search on life expectancy calculators, with focus on calculators that may be valuable for use in the Veterans Health Administration population receiving primary care.

With input from Operational Partners and Technical Expert Panel members, we developed the following Key Questions for the review:

Key Question 1: Between 2011 and 2016, have there been any additional reports of life expectancy calculators that may have sufficient predictive accuracy for use in adult primary care practice?

Key Question 2: Of the life expectancy calculators being reviewed have any external validation studies been published between 2011 and 2016? If yes, what population was studied and what was the predictive accuracy therein?

Key Question 3: What is the clinical use of the mortality prediction models (*aka* life expectancy calculators), and was there improvement in patient survival times, health-related quality of life, provider-patient communication, patient satisfaction and participation in clinical decisions, or lower healthcare utilization and costs?

METHODS

Data Sources and Searches

We searched MEDLINE (Ovid) from 2011 to May 2016 using title words for life expectancy, calculators or models, survival, mortality, death, and validation or calibration. The search was limited to English language and studies of humans middle-aged (45 plus years) and older. We also limited the search to relevant study designs.

Study Selection

The conceptual basis for this review was that a decision support tool would be electronically implemented to provide quantitative survival estimates to primary care providers who along with their patients would use the estimates when making healthcare decisions and thereby improve patient outcomes and healthcare value. We included studies of methods to estimate probabilities of survival or dying within a specified period of time. In addition, we were interested in studies of the use and effects of the resulting estimates.

We required that studies offer a method of estimation based on variables that would be generally available for primary care outpatients. Thus studies of estimates that require inpatient variables or patient surveys, studies of novel biomarkers, or studies of individual predictors, including those based on only age, sex, and race, were excluded. Studies that were limited to tests for associations such as relative risks, odds ratios, and hazard ratios were considered preliminary work and excluded.

Data Abstraction, Quality Assessment, Synthesis, and Strength of Evidence

Information on study sample and prediction model characteristics and performance were abstracted into evidence tables by one trained investigator and verified by the principal investigator.

We assessed quality of each study using selected items from the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS). Five criteria were selected: 1) predictor definition/measurement, 2) outcome definition/measurement, 3) independence of outcome and predictor assessments, 4) completeness of follow-up/predictor data, and 5) validation.

The evidence is narratively described without any formal meta-analysis. The strength of evidence for each key question is rated as high, low, or insufficient based on the number of quality studies and the consistency of the results.

RESULTS

Our literature search yielded 8,120 potentially eligible studies. Titles were reviewed by trained investigators and 509 studies were selected for abstract review. Each abstract was reviewed by 2 trained investigators and 51 studies were identified for full-text review. Full-text articles were reviewed by an investigator and the principal investigator. We included 10 studies from the literature search and an additional study identified from hand-searching studies included in the full-text review phase.

Key Question 1

We found 8 different mortality prediction models (in 11 reports). All had encouraging predictive accuracy. All but one model used patient's age and sex along with diagnostic codes extracted from electronic databases with or without additional variables. Models developed for mortality risk adjustment focused on mortality during the first year of follow-up while models intended for decisions about cancer screening had longer time horizons of 5 to 10 years. The best-characterized model based on VHA data, known as the Care Assessment Needs (CAN) score,

was not strictly a life-expectancy model as it focused on deaths without a preceding hospitalization rather than all deaths.

Overall, the 8 models were reasonably well-calibrated, reporting relatively small differences between observed and predicted mortality in arbitrary mortality risk groups. Thus, if a clinician used one of the models to place a patient in a mortality risk group, the average or median predicted mortality for the risk group could be used as an estimate. However, some individual predicted probabilities of survival may be substantially different from the average of the risk group. Furthermore, the differences in estimates between risk groups may not be sufficient to prompt clinicians or patients to make different decisions about screening or treatment.

The strength of evidence of the body of evidence for Key Question 1 was rated as high based on finding several studies of acceptable quality that reported a mortality prediction model based on similar predictor variables that had potentially useful predictive accuracy.

Key Question 2

One of the life expectancy models identified in Key Question 1 was in a sense validated in the older patients from the population used to generate the model and a sub-population of individuals with schizophrenia. Another model was in a sense validated in the older patients and a sub-population of heart failure patients. Calibration was not always as good in the sub-populations.

Not finding any true external validation studies of the prediction models in primary care practices, we rated the strength of evidence for Key Question 2 as insufficient.

Key Question 3

We found no studies that examined the effects of using the reviewed mortality prediction models in clinical care.

The strength of evidence for Key Question 3 was rated as insufficient.

DISCUSSION

Key Findings

- Between 2011 and 2016, 11 studies reported on 8 mortality prediction models; all but one of the included studies (a study from Japan) utilized data from large electronic databases.
- Models were developed for different purposes including development of individual risk scores to adjust for possible difference in mortality risk when comparing healthcare outcomes, to help primary care teams assess short-term risk of hospitalization or death without hospitalization, or to help healthcare providers judge whether patient will or won't die within a specified time pertinent to decisions about screening for cancer.
- C-statistics ranged from 0.77 to 0.90, indicating the models provide good discrimination of those who survived or died during the varying periods of follow-up. Few studies reported the sensitivity, specificity, or positive or negative predictive values for a proposed risk score threshold that would be used to determine which patients should or shouldn't be screened or treated.

- The prediction models were generally well-calibrated to the test samples with seemingly insubstantial differences between observed and predicted mortality across a range of risk groups.
- We found no true external validation studies of the reviewed mortality prediction models. None of the models have been externally validated for general primary care use.
- No studies meeting eligibility criteria for the review examined the impact of using one of the life expectancy calculators to improve healthcare decisions or outcomes.

Applicability

Several of the prediction models identified for this review could potentially be adapted for use in VA primary care practices. However, the discrimination and calibration of the models may vary in VA primary care practices that may have substantially different distributions of the predictors and/or mortality, and therefore would need to be evaluated. To adapt a mortality prediction model, the sources of data and definitions of predictor variables will need to be consistent across VA healthcare systems and over time. Centralized prediction models that are made widely available should be calibrated to and validated in several VA primary care patient populations and periodically checked and recalibrated if necessary. Although feasible, we found no studies to indicate whether making a reasonably accurate life expectancy calculator available to VA primary care providers or patients would influence their healthcare decisions or outcomes.

Research Gaps/Future Research

Although healthcare providers and guidelines make recommendations based, in part, on assessments of life expectancy, there is no widely accepted statistical tool for estimating patients' life expectancy or probability of survival, particularly for prolonged periods, for example, 10 years. Research on the clinical usefulness and impact of mortality prediction models on clinical decisions and outcomes is needed to guide further development and engender widespread acceptance. Analytical life expectancy predictions have to be demonstrated to be more accurate than clinicians' intuitive prognostic assessments, preferably using statistics that allow clinicians to compare prognostic errors. Additionally, they should use patient information that is readily and reliably available. Strong comparative evidence that using a quantitative prediction model can reduce practice variation and improve healthcare decisions and outcomes most likely will be needed to change current practices.

Conclusions

Life expectancy calculators based on readily available electronic data that have acceptable performance for estimating one-, 5-, and 10-year life expectancy in middle age to older adults are feasible. These calculators need to be validated for use in primary care practices. There are no data on the effect of using these life expectancy calculators on patient or provider decisions or outcomes. If a life expectancy calculator is made available, it remains to be determined whether primary care providers would use it or whether it would improve healthcare delivery, resource use, patient experiences, or outcomes.

EVIDENCE REPORT

INTRODUCTION

Life expectancy, an estimate of the number of remaining years of life a person has, is an important consideration for making clinical decisions in primary care. For example, colorectal cancer screening guidelines state that clinicians should only screen patients with an estimated life expectancy of at least 10 years because otherwise benefits of cancer detection are unlikely to outweigh the harms and costs. Referral to hospice care is often based on a life expectancy of less than 6 months. Implantable cardiac defibrillators are not indicated if the patient is not expected to live longer than one year.

Most currently available life expectancy calculators or life tables are based on a person's age, gender, and race. These calculators may not be widely used in clinical practice because clinicians usually consider other key factors such as the presence and severity of life-threatening diseases and functional status. Given the uncertainty inherent in formulating a prognosis and desire to avoid prognostic errors, clinical prognostic assessments are often qualitative, such as thinking a patient has a 'higher' risk of dying, and often are not shared with patients.¹ In contrast, survival prediction models typically incorporate a number of variables to calculate a quantitative estimate of the patient's probability of surviving or dying during a specified period of time.

This systematic review focused on identifying and evaluating reports of multivariable quantitative prediction models (*aka* calculators) of all-cause mortality published in 2011 and thereafter. Others have reviewed reports of predictive models for older patients from before 2011.^{2,3} These previous reviews listed a large number of prediction models that are available for primary care or population-based settings. However, the reviewers stated the evidence was insufficient to support their widespread clinical use. Of interest for this review were prediction models of all-cause mortality that would generally be applicable to most patients seen in primary care practices without off-putting effort by clinicians to ascertain the predictor variables and calculate the estimates. In addition, we were interested in reports that provided assessments of a proposed model's predictive accuracy, external validity, and ideally impact on clinical decision-making and patient outcomes. The ultimate goal is to identify and evaluate life expectancy calculators that primary care providers would be willing and able to use and share with their patients to improve participation and satisfaction with clinical decisions that are based, in part, on life expectancy. Ultimately, efforts to make a validated and accurate life expectancy calculator readily available to clinicians will need to demonstrate benefits in terms of improving healthcare outcomes and efficiency as well as patient experiences.

The Key Questions for this review and our approach to evaluating the pertinent evidence were as follows.

KQ1: Between 2011 and 2016, have there been any additional reports of life expectancy calculators that may have sufficient predictive accuracy for use in adult primary care practice?

Most mortality prediction models are not used to calculate a patient's life expectancy (number of years of life remaining on average) *per se*. Rather they estimate probabilities of surviving or

dying within a specified period of time. If all patients being studied were followed to the end of a time period of interest, for example, 5 years, then multivariable logistic regression models are typically used to estimate a patient's probability of dying before the end of the period. As a proxy for life expectancy, one might say if one's estimated probability of dying within 5 years is greater than 0.5 (50%), one's life expectancy is less than 5 years.

During development of mortality prediction models, the estimated probabilities of dying are often arbitrarily categorized into risk groups, and the predicted number of deaths in each risk group is compared to the observed number. If the differences are small, the prediction model was well-calibrated to the studied sample. Primary care providers may be more willing to use a well-calibrated model, especially if patients placed in particular well-calibrated risk groups would be treated differently (presumably because the intervention has a different likelihood of benefiting or harming patients in a particular risk group). Ideally the range of individual risk estimates within a risk group would be narrow, making it easier to apply the risk group's average probability of dying to individuals within the group.

If a study couldn't follow all subjects for the entire time period of interest, the varying follow-up times can still be used to estimate survival (or mortality) curves. Survival curves have time on the x-axis and the estimated cumulative proportion surviving on the y-axis. The median survival time is the time when a survival curve reaches 0.5 on the y-axis. The median survival time can also be used as a proxy for life expectancy (an average) because half of the group lived longer and half shorter than the median survival time. To estimate life expectancy, one would have to fit a parametric equation to the survival curve and extrapolate it to cover the period of interest or until the curve reaches zero (all people in the cohort are deceased).^{4,5} Cox proportional hazards regression models are often used to relate multiple predictor variables to the survival times. One can use a fitted Cox regression model to estimate the probability of surviving (or dying) at a specified time or to estimate a survival (or mortality) curve using a patient's values of the predictor variables.

A C-statistic is commonly reported to help evaluate the ability of a mortality prediction model to identify (discriminate) the patients who did or didn't die within the period of follow-up. It is a measure of concordance or correlation (hence, the name "C-statistic") between observed and estimated of survival probabilities or times. Thus, C-statistics measure the ability of a model to rank patients according to their risk but do not assess the ability of a model to assign accurate probabilities of surviving or the model's calibration. A C-statistic equal to 0.5 indicates the model did not discriminate those who survived or died during the period of follow-up any better than flipping a coin.

If a model's calculated risks are categorized using a particular cut-off to predict whether a patient will or will not die within a specified period of time, then the sensitivity (the proportion of all deaths that had risk scores exceeding the cut-off), and the specificity (the proportion of survivors that had risk scores less than the cut-off) can be estimated for the cut-off. Whether any cut-offs have a sufficient sensitivity and/or specificity for clinicians to use needs to be determined. The likelihood of finding a cut-off that has both a high sensitivity and specificity increases as the value of the C-statistic approaches 1.0. Given a proposed cut-off for making prognostic predictions, the positive predictive value (the proportion of patients above the cut-off that would be predicted to die and do) and the negative predictive value (the proportion of patients below the cut-off that would be predicted to survive that do) can be estimated. Primary care providers

may require a very high positive and/or negative predictive value – that is, few prediction errors – before they're willing to use a quantitative model to predict whether a patient will survive.

KQ2: Of the life expectancy calculators being reviewed, have any external validation studies been published between 2011 and 2016? If yes, what population was studied and what was the predictive accuracy therein?

Validation refers to testing a prediction model in a new sample of patients that was not used to develop the model. Ideally, a mortality prediction model should be validated in the patient population in which it will be used or a very similar patient population. Often validation is done by randomly splitting a sample of patients into one or more subsamples and using one subsample to develop the model and another subsample to validate the model. However, this approach may be overly optimistic in regards to future predictive performance because the distributions of predictor variables and mortality tend to be similar in randomly split samples. Validation studies that used different patients and time periods were considered to be more informative.

Before validating the predictive accuracy in a new sample of patients, it is appropriate to check the calibration of the prediction model in the new sample and perhaps recalibrate it before examining the predictive performance. Some validation studies may simply re-estimate the regression coefficients for the predictor variables although this practice is discouraged because the new estimates may be less precise if the external validation sample is small and less generalizable than re-calibrated estimates.

KQ3: What is the clinical use of the mortality prediction models (*aka* life expectancy calculators), and was there improvement in patient survival times, health-related quality of life, provider-patient communication, patient satisfaction and participation in clinical decisions, or lower healthcare utilization and costs?

To address this key question, we were particularly interested in finding randomized controlled trials that compared clinical use of a mortality prediction model to usual care without the prediction model.

PICOTS

Population: Middle- to older-aged adults (age 45 years and older) that were not in a hospital at the start of follow-up. If the study included a younger adult age range, it was included especially if predictive accuracy was evaluated in a more elderly subgroup.

Intervention: Calculation of life expectancy (or a proxy) using predictor variables that are generally available for outpatients being seen by primary care providers. Less generalizable disease-specific mortality prediction models or those that incorporated disease-specific or novel predictor variables were excluded. Models that predicted inpatient mortality or used a number of inpatient-specific predictor variables that would only be available for hospitalized patients were also excluded. Models that incorporated inpatient diagnosis codes as well as outpatient diagnosis codes were included.

Comparator: None

Outcomes

KQ1, KQ2: Predictive accuracy in a validation sample. See text under KQ1 and KQ2 for a detailed explication.

KQ3: Clinical use of a prediction model, improvement in patients' survival times, health-related quality of life, provider-patient communication, patient satisfaction and participation in clinical decisions, and healthcare utilization and costs when prediction model was made available to primary care providers.

Timing: Any follow-up time period that may be pertinent to a clinical decision.

Setting: Outpatient primary care.

METHODS

TOPIC DEVELOPMENT

The topic was nominated by Dr. David MacPherson (Chief Medical Officer, VISN4). Drs. Stephan Fihn (Office of Analytics and Business Intelligence) and Joe Francis (Clinical Analytics and Reporting) were consulted and agreed to be operational partners for the review. The key questions and scope of the review were developed with input from the operational partners and a Technical Expert Panel.

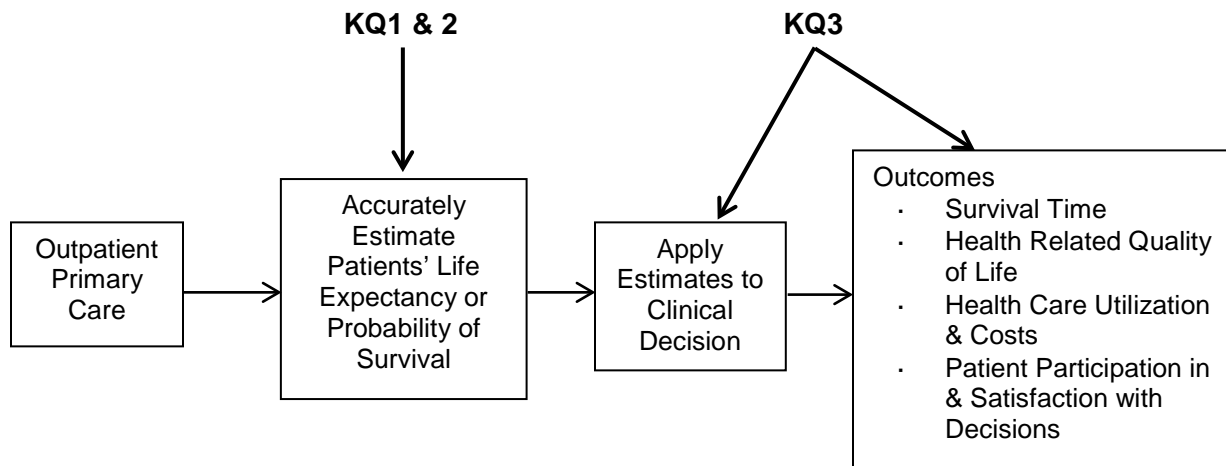
SEARCH STRATEGY

We searched MEDLINE (Ovid) from 2011 to May 2016 using title words for life expectancy, calculators or models, survival, mortality, death, and validation or calibration. The search was limited to English language and studies of humans middle-aged (45 plus years) and older. We also limited the search to relevant study designs. The full search strategy is presented in Appendix A.

STUDY SELECTION

The following conceptual framework (Figure 1) guided study selection.

Figure 1. Conceptual Framework



The conceptual basis for this review was that a decision support tool would be electronically implemented to provide quantitative survival estimates to primary care providers who along with their patients would use the estimates when making healthcare decisions and hopefully improve patient outcomes and the efficiency of healthcare. For example, healthcare providers might place patients into risk groups that would be treated differently or they might somehow share the estimates with patients. Either individual or risk group estimates of life expectancy or probabilities of survival or dying within a specified period of time were of interest.

The methods of estimation had to be generally applicable to primary care patient populations. Hence a large numbers of disease- or cause-specific mortality prediction models were excluded. Included studies had to offer a method of estimation based on variables that would be generally

available for outpatients. Estimates that require inpatient variables that would only be available for a minority of primary care patients were excluded. Likewise studies of novel biomarkers or individual predictors, including those based on only age, sex, and race, were excluded. Reports of predictive values (percentage of predictions that were correct), including group survival curves, observed versus predicted survival/mortality (*aka* calibration), and sensitivity or specificity of predictions were considered to be most pertinent to clinical decision-makers and thus this review. Studies that were limited to tests for associations such as relative risks, odds ratios, and hazard ratios were considered to be too preliminary for inclusion in this review.

Studies of the clinical use of survival estimates were of particular interest. For example, estimates might be used to help providers and patients make a specified decision about prevention or treatment. Estimates might be used to help healthcare providers prioritize their patients for medical evaluation or services according to the expected net benefit based, in part, on life expectancy. Clinical trials might have used the prediction model to help select or identify patients for whom a treatment is most likely to be of benefit or least likely to harm.

The effects of applying prognostic estimates on patients' survival and/or health-related quality of life are the ultimate outcomes of interest. In addition, we sought studies of the effects of using survival estimates on healthcare utilization, costs, and patient satisfaction or participation in healthcare decisions.

DATA ABSTRACTION

Data were abstracted into evidence tables by one trained investigator and verified by the principal investigator. We abstracted information on study characteristics (country, study dates, source of data, participants, mortality) and model characteristics and performance (intended use of model, predictors, timing of predictor and survival assessments, modelling method, methods used to select predictors, mortality risk groups, and predictive performance).

QUALITY ASSESSMENT

Quality of individual studies was assessed using selected items from the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS).⁶ Five criteria were selected: 1) predictor definition/measurement, 2) outcome definition/measurement, 3) independence of outcome and predictor assessments, 4) completeness of follow-up/predictor data, and 5) validation.

DATA SYNTHESIS

The evidence is narratively described without any formal meta-analysis.

RATING THE STRENGTH OF EVIDENCE

The strength of evidence for each key question is rated as high, low, or insufficient based on the number of quality studies and the consistency of the results.

PEER REVIEW

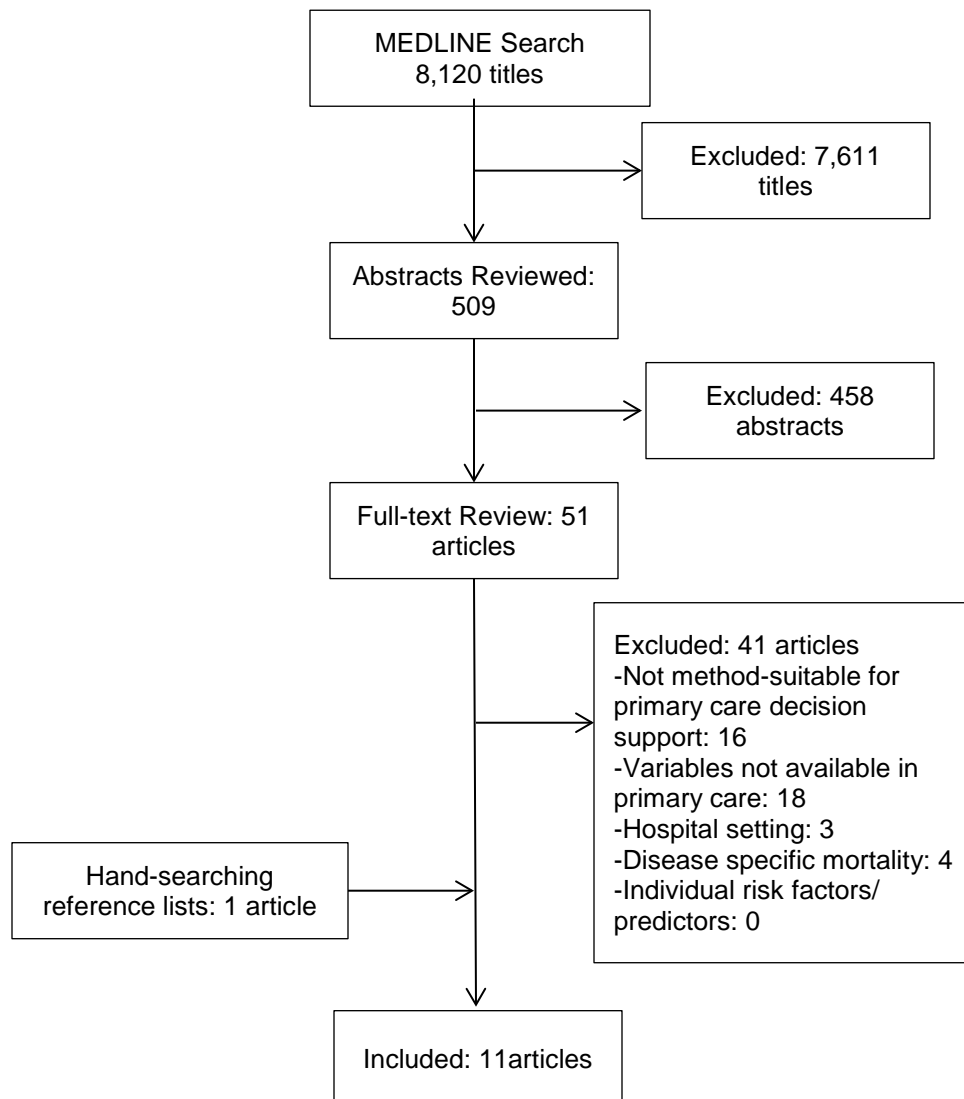
A draft version of this report was reviewed by content experts as well as clinical leadership. Reviewer's comments and our responses are presented in Appendix B and the report was modified accordingly.

RESULTS

LITERATURE FLOW

Our literature search yielded 8,120 titles (Figure 2). Titles were reviewed by trained investigators and 509 studies were selected for abstract review. Each abstract was reviewed by 2 trained investigators and 51 studies were identified for full-text review. Full-text articles were reviewed by an investigator and the Principal Investigator. We included 10 studies from the literature search and an additional study identified from hand-searching studies included in the full-text review phase.

Figure 2. Literature Flow Chart



Reviewed Studies

Table 1 summarizes the 11 eligible studies of 8 different mortality prediction models (see Appendix C, Tables 1 & 2 for more details). Six studies were done in the US.⁷⁻¹² Three were based on large Veterans Health Administration (VHA) electronic databases.^{8,11,12} Others utilized Medicare data^{7,9} or local electronic health records.¹⁰ Four studies were from Canada using electronic administrative data representing the entire provinces of Ontario¹³⁻¹⁵ or Saskatchewan.¹⁶ One study from Japan used data collected by other studies combined with an extensive search for deaths.¹⁷ The median age of the Canadian provincial cohorts was in the mid-forties range, and over 60 in the other studies. All but the VA studies had approximately equal or greater inclusion of women and men. Only 2 studies examined race as a potential predictor.^{8,10}

Table 1. Study Population Characteristics^a

Study	Location Year(s)	Population	Primary Data Source	Subjects
Austin 2011 ¹⁵	Ontario, Canada 2007-08	general adults	provincial health plan	n=10,498,413 median age 46 female 51%
Austin 2011 ¹⁴	Ontario, Canada 2007-08	general adults	provincial health plan	n=10,498,413 median age 46 female 51%
Austin 2012 ¹³	Ontario, Canada 2007-08	adults with schizophrenia	provincial health plan	n= 94,466 median age 47 female 54%
Gagne 2011 ⁷	PA, NJ, USA 2004-05	low income Medicare	Medicare & Pharmacy Assistance Programs	n=120,679 mean age 80 female 83%
Quail 2011 ¹⁶	Saskatchewan, Canada 2001-02	general adults	provincial health plan	n=662,423 median age 48 female 51%
Stefos 2012 ⁸	USA 2007-08	VHA healthcare	VHA electronic records	n=4,774,000 mean age 62 female 6%
Mathias 2013 ¹⁰	Chicago, USA 2003-05	multispecialty group practice	electronic health records (EPIC)	n=7,463 mean age 62 female 60%
Tan 2013 ⁹	USA 1999-09	2000 Medicare enrollees	Medicare	n=1,137,311 mean age 76 female 60%
Ogata 2013 ¹⁷	Kyushu Island, Japan 1999-09	representative cohorts from other studies	study data plus extensive search for death records	n=2021 mean age 63 female 59%
Wang 2013 ^{12b}	USA 2010-12	VHA healthcare	VHA electronic records	n=4,598,408 mean age 64 female 6%
Wang 2012 ^{11b}	USA 2009-10	VHA healthcare with heart failure	VHA electronic records	n=198,640 mean age 73 female 2%

^a Shaded rows are VHA studies

^b VA Care Assessment Needs (CAN) model

As summarized in Table 2, 4 of the mortality prediction models presented in 6 articles were developed to calculate individual risk scores to adjust for possible differences in mortality risk when comparing healthcare outcomes in different groups of patients.^{7,8,13-16} One VHA model was developed to help primary care teams assess the short-term risk of hospitalization or death without hospitalization.^{11,12} Three prediction models were developed to help healthcare providers judge whether patients would or would not die within a specified time.^{9,10,17} All of the prediction models incorporated the patients' age and sex and all except one¹⁷ included selected diagnoses. All of the predicted models that incorporated diagnosis codes used a method to group ICD-9 or ICD-10 diagnosis codes into predictor variables, except Mathias 2013.¹⁰ A varying number included other demographic and clinical assessments such as vital signs, laboratory results, and prescription medications as predictor variables.

Models developed for mortality risk adjustment focused on mortality during the first year of follow-up, when the overall mortality ranged from < 1% to 7.5% depending on the types of patients selected for the study. Other models focused on whether patients would be expected (predicted) to live for time horizons of 5 or 10 years.^{9,10,17} These longer time horizons had higher mortality. To assess care needs, Wang et al modeled deaths that occurred without a hospitalization within one year rather than total all-cause mortality, but did so using VHA data.^{11,12} Presumably the same data could be utilized to model total mortality.

Table 2. Description of Prediction Models^a

Study	Intended Use	Types of Predictors	Mortality being Predicted (mortality%)
Austin 2011 ¹⁵	Statistical risk adjustment	Patients' age, sex, and diagnoses ^b recorded in previous 2 years	Death within 1 year (0.8%) Subgroup age > 65 (NR)
Austin 2011 ¹⁴	Statistical risk adjustment	Patients' age, sex, and diagnoses ^b recorded in previous 2 years	Death within 1 year (0.8%)
Austin 2012 ¹³	Statistical risk adjustment	Patients' age, sex, and diagnoses ^b recorded in previous 2 years	Death within 1 year (2.0%)
Gagne 2011 ⁷	Statistical risk adjustment	Patients' age, sex, and diagnoses ^c recorded in previous year	Death within 1 year (7.5%)
Quail 2011 ¹⁶	Statistical risk adjustment	Patients' age, sex, income, region, and diagnoses ^c recorded in previous year	Death within 1 year (1.3%) Subgroup age > 65 (5.1)
Stefos 2012 ⁸	Statistical risk adjustment	Selected patient demographics and diagnoses, ^d including chronic disease registries, in previous year	Death within 1 year (5.5%)
Mathias 2013 ¹⁰	Health care decision	Selected patient demographics, diagnoses, vital signs, laboratory test results, medications, and prior healthcare utilization in prior 1-2 years	Death within 5 years (11%)
Tan 2013 ⁹	Health care decision	Patients' age, sex, and diagnoses, ^c and healthcare utilization recorded in previous year	Death within 10 years (women 51%; men 57%)
Ogata 2013 ¹⁷	Health care decision	Cardiovascular risk: factors age, sex, smoking, blood pressure, cholesterol, glycated hemoglobin	Death within 10 years (site 1 13%; site 2 28%)
Wang 2013 ^{12e}	Identify patients at higher risk of hospitalization taking competing risk of death into account	Selected patient demographics, diagnoses, vital signs, laboratory test results, medications, and prior healthcare utilization in previous year	Death without hospitalization within 90 days (0.7%) & 1-year (2.8%)
Wang 2012 ^{11e}	Identify patients at higher risk of hospitalization taking competing risk of death into account	Selected patient demographics, diagnoses, ^{b,d} vital signs, laboratory test results, medications, and prior healthcare utilization in previous year	Death without hospitalization within 30 days (0.9%) & 1-year (7.1%)

^a Shaded rows are VHA studies

^b Diagnosis codes grouped using Aggregated Diagnosis Group software

^c Diagnosis codes grouped into Charlson & Elixhauser categories

^d Diagnosis codes grouped using Hierarchical Coexisting Conditions

^e VA Care Assessment Needs (CAN) model

NR = not reported

Study Quality

Most studies were judged to have acceptable quality based on meeting at least 4 of the 5 criteria (Table 3). The extent of missing predictor variables often was not reported, presumably because a relatively small proportion of subjects with missing predictor values were excluded from model development and validation. The Ogata study that utilized data from 2 previous cohort studies had extensive amounts of missing data predictor values for one study site and the completeness of the post hoc searches for dates of death in the community wasn't clear.¹⁷

Table 3. Quality of Included Studies^a

Study author, year	Predictor definition or measurement same for deceased/survivors	Outcome definition or measurement same for deceased/survivors	Outcome assessed independent of predictors (eg, blinded)	Incomplete follow-up or missing predictor data (%; handling of)	Method of validation	Overall quality
Austin, 2011 ¹⁵	Yes	Yes	Yes	NR	Assessed over fitting by bootstrap methods	Acceptable
Austin, 2011 ¹⁴	Yes	Yes	Yes	NR	Assessed over fitting by bootstrap methods	Acceptable
Austin, 2012 ¹³	Yes	Yes	Yes	NR	Application of previously developed model for general adult population	Acceptable
Gagne, 2011 ⁷	Yes	Yes	NR	NR	External sample	Acceptable
Quail, 2011 ¹⁶	Yes	Yes	Yes	Missing income imputed	External sample & time	Acceptable
Stefos, 2012 ⁸	Yes	Yes	Yes	NR	Copas test for overfitting (repeated, split sample, cross-validation design)	Acceptable
Mathias, 2013 ¹⁰	Yes	Yes	Yes	NR	10-fold cross validation	Acceptable
Tan, 2013 ⁹	Yes	Yes	Yes	NR	Random split sample	Acceptable
Ogata, 2013 ¹⁷	Yes	Yes	NR	3.6% missing data/lost at site 1; 80% excluded at site 2 due to missing data/lost or < 10 years follow-up	1) Split sample (random) from first site and 2) second site only served as validation site	Unacceptable
Wang, 2013 ^{12b}	Yes	Yes	Yes	Yes	Randomly split sample	Acceptable
Wang, 2012 ^{11b}	Yes	Yes	Yes	Yes	Randomly split sample	Acceptable

^a Shaded rows are VHA studies^b VA Care Assessment Needs (CAN) model

NR = not clearly reported

KEY QUESTION 1: Between 2011 and 2016, have there been any additional reports of life expectancy calculators that may have sufficient predictive accuracy for use in adult primary care practice?

Summary of Findings

Most studies reported a range of estimated probabilities of dying that included some values greater than 0.5 (50%) that could be interpreted as having a life expectancy less than the model's time horizon (Table 4). Wang et al modeled death without a hospitalization, thus their model cannot be used as is to estimate life expectancy.^{11,12} However, as noted previously, most likely their prediction model could be adapted to do so.

The reported C-statistics ranged from 0.77⁹ to over 0.90^{15,16} indicating that the models discriminated those who survived or died by rank ordering subjects according to their risk of dying, but not perfectly. The latter 2 studies achieved these high levels of discrimination using the patient's age and gender in combination with diagnostic codes extracted from large provincial health plan databases to predict 1-year mortality. These very high levels of discrimination most likely were related to inclusion of a large proportion of younger adults that would be less likely to die than older adults. In contrast, Gagne et al, who used similar predictors extracted from state health plan databases, only had a C-statistic of 0.79 in a more elderly sample of patients with higher mortality at one year.⁷ Tan et al also used similar predictors extracted from Medicare data; however their lower C-statistic is for deaths that occurred over a much longer 10-year time span (these are inherently more difficult to estimate than a shorter time span due to interim changes in predictors and health).⁹ Using Tan et al's 10-year mortality prediction model, over 80% of the subjects that had a predicted 10-year probability of dying that was greater than 0.75 (75%) died within this time interval whereas approximately 65% with a 10-year probability of dying between 0.50 to 0.74 died. The positive and negative predictive values were 63% and 92% using a cut-off for Mathias et al's predicted probability of dying within 5 years equal to 0.5 (50%).¹⁰ Other reports did not focus on any particular risk score threshold to predict survival and therefore did not estimate sensitivity, specificity, or predictive values.

The differences between observed and predicted mortality in subjects that were grouped according to their estimated probabilities of dying were mostly small, that is, the prediction models were well-calibrated to the test sample. Thus, if a clinician used the prediction model to place a patient in a mortality risk group, the observed mortality in the group most likely would be similar to the risk group's average or median predicted mortality. If a risk group's individual estimated probabilities of dying within the specified follow-up time are all either greater than or less than 0.5, as a proxy one might say their life expectancy is either less than or greater than the specified follow-up time, respectively.

Table 4. Predictive Performance of Models under Review^a

Study	Estimated Probabilities of Dying	Discrimination	Calibration†																						
Austin 2011 ¹⁵	0.00 to 0.90	C-statistic: 0.92	Differences between observed and predicted mortality <1% in all 100 risk categories except top 3 where differences < 3%																						
Austin 2011b ^{b14}	0.00 to 0.90	C-statistic: 0.92	Differences between observed and predicted mortality ~1% in all 20 risk categories																						
Gagne 2011 ⁷	Mean of 1 st decile 0.03 to 10 th decile 0.55	C-statistic: 0.79	Under predicted by ~ 3% in 5 th & 6 th deciles with 20 -25% mortality, over predicted by ~ 10% in 10 th decile with ~ 45% mortality, otherwise observed and predicted mortality % within ~1%																						
Quail 2011 ¹⁶	NR	C-statistic with Charlson comorbidities: 0.90 C-statistic with Elixhauser comorbidities: 0.91	NR																						
Stefos 2012 ⁸	Mean of 1 st decile 0.025 to 10 th decile 0.94	C-statistic: 0.86	Ratio of observed to predicted number of deaths by predicted risk decile <table border="1"> <thead> <tr> <th>Predicted</th> <th>O/P ratio</th> </tr> </thead> <tbody> <tr><td>2.5%</td><td>0.94</td></tr> <tr><td>13%</td><td>1.12</td></tr> <tr><td>24%</td><td>1.12</td></tr> <tr><td>34%</td><td>1.06</td></tr> <tr><td>45%</td><td>0.98</td></tr> <tr><td>55%</td><td>0.95</td></tr> <tr><td>65%</td><td>0.91</td></tr> <tr><td>75%</td><td>0.87</td></tr> <tr><td>85%</td><td>0.87</td></tr> <tr><td>94%</td><td>0.86</td></tr> </tbody> </table>	Predicted	O/P ratio	2.5%	0.94	13%	1.12	24%	1.12	34%	1.06	45%	0.98	55%	0.95	65%	0.91	75%	0.87	85%	0.87	94%	0.86
Predicted	O/P ratio																								
2.5%	0.94																								
13%	1.12																								
24%	1.12																								
34%	1.06																								
45%	0.98																								
55%	0.95																								
65%	0.91																								
75%	0.87																								
85%	0.87																								
94%	0.86																								
Mathias 2013 ¹⁰	Mean of 1 st decile 0.036 to 10 th decile 0.55	C-statistic: 0.86 Sensitivity/specificity using cut-off for predicted probability of dying within 5 years = 0.5 (50%): 31%/98% Positive/negative predictive value using cut-off for predicted probability of dying within 5 years = 0.5 (50%): 63%/92%	Differences between observed and predicted mortality <3% across all deciles of predicted risk																						
Tan 2013 ⁹	1 st quartile < 0.25 to 4 th quartile > 0.75	C-statistic at 10 years: Women: 0.79 Men: 0.7 Positive predictive value for LE < 10 years for women/men: ~75%/75%	Observed mortality within quartiles of predicted probabilities for women/men: 1st quartile: 17%/20% 2nd quartile: 35%/35% 3rd quartile: 65%/64% 4th quartile: 88%/90%																						

Ogata 2013 ¹⁷	Predicted to die in 10 years: Site 1: 36/365 (9.9%) Site 2: 35/170 (20.6%)	C-statistic Site 1: 0.83 Site 2: 0.85	Difference between observed and predicted mortality: Site 1: 3.5% Site 2: 7.6%
Wang 2013 ^{12c,d}	1-year mean of 1 st decile < 0.01 to 10 th decile ~0.14	C-statistic: 0.85	Differences between observed and predicted in each decile of predicted probabilities: negligible Observed deaths if in upper 5% of 1-year predicted probabilities of death without hospitalization: 19.4%

^a Shaded rows are VHA studies

^b Same prediction model as Austin 2011¹⁵ except regression coefficients converted to a single weighted Mortality Risk Score

^c Only predicting deaths without hospitalization, not total deaths

^d VA Care Assessment Needs (CAN) model

NR = not reported

Conclusion for Key Question #1

Since 2011 several life expectancy calculators have been reported that may have sufficient predictive accuracy for general use in adult primary care practice. Accurate proxy life expectancy calculators can be developed using age, gender, and diagnosis codes and perhaps a few other select variables obtained from administrative or electronic medical records. However, all of the models under review would require some adaptation before being made available to VA primary care providers. Few studies reported the sensitivity, specificity, or positive or negative predictive values for a risk threshold proposed for a clinical decision such as whether a patient should or shouldn't be screened or treated.

Rating for the Strength of Evidence for KQ1: High

Several prediction models utilizing variables that are generally available in electronic medical records had consistently good predictive accuracy.

KEY QUESTION 2: Of the life expectancy calculators being reviewed, have any external validation studies been published between 2011 and 2016? If yes, what population was studied and what was the predictive accuracy therein?

Summary of Findings

Our review didn't find any true validation studies. We found 2 separate quasi-validation reports^{11,13} and 2 elderly subgroup analyses.^{12,15} The predictive performance in these patient subgroups is summarized in Table 5. Regression coefficients for the predictor variables were re-estimated for the 2 elderly subgroup analyses^{12,15} and the Wang et al study of patients with heart failure prior to examining the predictive accuracy in the subpopulations.¹¹ Compared to the mortality prediction models for general adult patient populations, both the Austin and Wang models had poorer discrimination of survivors and decedents in the elderly and schizophrenic or heart failure subpopulations. Calibration of the Austin model was poorer in the schizophrenic subpopulation whereas after recalibration the calibration of the Wang model was maintained in the elderly and heart failure subpopulations.

Table 5. Validation of Mortality Prediction Models^a

Study	Population Subgroup	Number Mortality	Discrimination	Calibration
Austin 2011 ¹⁵	Age ≥ 65 years	n=NR 1-year NR	C-statistic: 0.81	Differences between observed and predicted mortality NR
Austin 2012 ¹³	Schizophrenia	n=94,466 1-year 2.0%	C-statistic: 0.84	Differences between observed and predicted mortality NR, however calibration plots indicated significantly worse calibration
Wang 2013 ^{12b,c}	Age > 65 years	n=2,129,063 1-year ~4.7%	C-statistic: 0.80	Small differences between observed and predicted in each decile of predicted probabilities
Wang 2012 ^{11b,c}	Heart failure	n=198,640 1-year 7.1%	C-statistic: 0.76	Differences between observed and predicted mortality % in each decile of predicted probabilities were negligible

^a Shaded rows are VHA studies

^b Only predicting deaths without hospitalization, not total deaths

^c VA Care Assessment Needs (CAN) model

NR = not reported

Conclusion for Key Question #2

Discrimination of the deceased versus surviving patients was reduced in these more homogenous subpopulations with higher mortality than the overall general populations. Calibration of the prediction models was not always as good in the subpopulations despite recalibration. The calibration of a mortality prediction model should be examined in a representative sample of the

primary care practice in which it will be used, and the predictive accuracy should be validated using time horizons that are commensurate with the clinical decisions it will support.

Rating for the Strength of Evidence for KQ2: Insufficient

None of the 4 studies were true validation studies that examined the predictive performance in external samples of primary care practices.

KEY QUESTION 3: What is the clinical use of the mortality prediction models (*aka* life expectancy calculators), and was there improvement in patient survival times, health-related quality of life, provider-patient communication, patient satisfaction and participation in clinical decisions, or lower healthcare utilization and costs?

Summary of Findings

We found no studies that examined the effects of using the reviewed mortality prediction models in clinical care.

We found 2 reports that alluded to use of other life expectancy calculators. One study retrospectively evaluated the potential impact of a previously developed 4-year life expectancy calculator (Health and Retirement Study Mortality Risk Index¹⁸).¹⁹ A retrospective patient cohort of 8,090 that was age 65 and older at the time of their last visit to a primary care practice in 2003-04 was studied. The prediction model identified 1,241 of the 8,090 (15%) as having a 40 to 70% probability of dying within 4 years. Their observed mortality was 670/1,241 (54%); thus, the authors concluded it would have been reasonable for these patients to forego a screening colonoscopy due to their limited life expectancy. Whether the primary care providers and patients would have come to the same conclusion with or without the life expectancy calculator was not determined.

Clinicians in the United Kingdom developed a 3-year mortality risk calculator for elderly patients with breast cancer to help determine whether they should undergo surgical (if expected to live at least 3 years) or medical therapy.²⁰ Their disease-specific prediction model (not included in this review) incorporated the American Society of Anesthesiologists (ASA) grade, age, ethnicity, bilateral or previous contralateral breast cancer, and nurse administered questionnaires to obtain the Charlson comorbidity score, Mini Mental State Examination score, Geriatric Depression Scale score, and Barthel and Instrumental Indexes of Activities of Daily Living. The prediction model was used in a Co-morbidity Clinic for patients who were potentially unfit or declined standard treatment for breast cancer, and was reportedly useful for shared decision-making about treatment. The report did not state exactly how the risk model was employed or communicated to patients or actually study its impact on the treatment decision.

Conclusion for Key Question #3

We found no studies of the effects of the reviewed mortality prediction models on clinical decisions and outcomes. Clinical outcome studies of evaluating life expectancy calculator for clinical decision-making are needed to support their implementation by healthcare organizations and use by clinicians and patients. Studies should focus on a well-defined clinical decision in the applicable population, for example, decisions about cancer screening based on 10 year life expectancy or palliative care based on life expectancy less than 6-12 months.

Rating for the Strength of Evidence for KQ 3: Insufficient

SUMMARY AND DISCUSSION

KEY FINDINGS

- Between 2011 and 2016, 11 studies reported on 8 mortality prediction models; all but one of the included studies (a study from Japan) utilized data from large electronic databases.
- Models were developed for different purposes, including development of individual risk scores to adjust for possible difference in mortality risk when comparing healthcare outcomes, to help primary care teams assess short-term risk of hospitalization or death without hospitalization, or help healthcare providers judge whether a patient will or won't die within a specified time pertinent to decisions about screening for cancer.
- C-statistics ranged from 0.77 to 0.90, indicating the models provide good discrimination of those who survived or died during the varying periods of follow-up. Few studies reported the sensitivity, specificity, or positive or negative predictive values for a proposed risk score threshold that would be used to determine which patients should or shouldn't be screened or treated.
- The prediction models were generally well-calibrated to the test samples with seemingly insubstantial differences between observed and predicted mortality across a range of risk groups.
- We found no true external validation studies of the reviewed mortality prediction models. None of the models have been externally validated for general primary care use.
- No studies meeting eligibility criteria for the review examined the impact of using one of the life expectancy calculators to improve on healthcare decisions or outcomes.

LIMITATIONS

None of the life expectancy calculators under review were developed or validated specifically to estimate life expectancy (average survival times). The mortality prediction models can provide individual or average estimates of the probability of surviving for a specific period for a risk group that may be used as proxies for life expectancy. The range of individual estimates of survival probabilities within risk groups may be an important consideration when using the group average to estimate life expectancy, but often was not reported. Furthermore, the uncertainty inherent in personalized estimates may preclude their use, especially if a prediction model isn't based on a large number of patients that were followed for periods of clinical interest such as 5 or 10 years. The best way to communicate model predictions to providers and patients was not addressed. A number of disease-specific prediction models that were excluded from this review may be better-suited to disease-specific clinical decisions.

APPLICABILITY OF FINDINGS TO THE VA POPULATION

Several of the prediction models listed in the KQ1 section seemingly could be adapted for use in VA primary care practices. The VA model of Wang et al (*aka* the Care Assessment Needs or CAN score) most likely could be modified to estimate the total mortality risk for time periods that are most relevant for specified clinical decisions.¹² The work by Stefos et al supports the notion that the VA electronic records can be used to produce well-calibrated proxy estimates of life expectancy.⁸ Several other studies support the impression that models based on patient demographics, the presence of life-threatening conditions, and perhaps history of healthcare utilization and other readily available variables can provide mortality estimates that have good, but less than perfect, predictive accuracy. The small numbers of quasi-validation studies suggest that the discrimination and calibration may vary when externally validated in VA primary care practices that may have substantially different distributions of the predictors and/or mortality. To adapt a mortality prediction model, the sources of data and definitions of predictor variables will need to be consistent across VA healthcare systems and over time. Centralized prediction models that are made widely available should be calibrated to and validated in several VA primary care patient populations and periodically checked for accuracy and recalibrated if necessary. The VA should consider models of survival curves that would provide estimates of life expectancy in addition to probabilities of survival.⁴

Although feasible, we found no studies to indicate whether making a reasonably accurate life expectancy calculator available to VA primary care providers or patients would influence their healthcare decisions or outcomes. A recent VA study by Tang et al suggested that having limited life expectancy as estimated by the investigators based on patients' age and Charlson comorbidity count was associated with an appropriately reduced rate of prostate-specific antigen screening for prostate cancer (39% when the estimated life expectancy was less than 5 years versus 77% with an estimated life expectancy of 10 or more years).²¹ Unfortunately, this study did not ascertain whether or how the clinicians estimated life expectancy or whether a VA effort to provide a validated quantitative calculator would influence their prognostic assessment or screening decisions.

RESEARCH GAPS/FUTURE RESEARCH

Although healthcare providers and guidelines make recommendations based, in part, on assessments of life expectancy, there is no widely accepted statistical tool for estimating patients' life expectancy particularly for prolonged periods of 10 years. Research on the clinical usefulness and impact of life expectancy calculators on clinical decisions and outcomes is needed to guide further development and engender widespread acceptance. Hundreds of statistical prognostic tools have been developed, but are not being used in practice.^{22,23} Readily available statistical life expectancy estimates require both clinical face validity and proven accuracy for diverse decision-relevant risk groups and periods of time.²⁴ Analytical life expectancy predictions have to be demonstrated to be more accurate than clinicians' intuitive prognostic assessments preferably using statistics that allow clinicians to compare prognostic errors. Additionally, they should use patient information that is readily and reliably available. Whether use of a standardized life expectancy calculator would reduce undesirable practice variation in regards to decisions involving estimates of life expectancy needs to be determined. Strong comparative evidence that using a quantitative prediction model can improve healthcare decisions and outcomes most likely will be needed to change current practices.

CONCLUSIONS

Life expectancy calculators based on readily available electronic data that have acceptable performance for estimating one-, 5-, and 10-year life expectancy in middle age to older adults are feasible. These calculators need to be validated for use in primary care practices. There are no data on the effect of using these life expectancy calculators on patient or provider decisions or outcomes. If a life expectancy calculator is made available, it remains to be determined whether primary care providers would use it or whether it would improve healthcare delivery, resource use, patient experiences, or outcomes.

REFERENCES

1. Schoenborn NL, Bowman TL, Cayea D, Pollack CE, Feeser S, Boyd C. Primary care practitioners' views on incorporating long-term prognosis in the care of older adults. *JAMA Intern Med.* 2016;176(5):671-678.
2. Yourman L, Lee S, Schonberg M, Widera E, Smith A. Prognostic indices for older adults: a systematic review. *JAMA.* 2012;307(2):182-192.
3. Minne L, Ludikhuizen J, de Rooij SE, Abu-Hanna A. Characterizing predictive models of mortality for older adults and their validation for use in clinical practice. *J Am Geriatr Soc.* 2011;59(6):1110-1115.
4. Lee SJ, Boscardin J, Kirby KA, Covinsky KE. Individualizing life expectancy estimates for older adults using the Gompertz Law of Human Mortality. *PLOS One.* 2014;9(9):e108540.
5. Baade PD, Youlten DR, Andersson TML, et al. Estimating the change in life expectancy after a diagnosis of cancer among the Australian population. *BMJ Open.* 2015;5:e006740.
6. Moons K, de Groot J, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLOS Med.* 2014;11(10):e1001744.
7. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol.* 2011;64(7):749-759.
8. Stefanos T, Lehner L, Render M, Moran E, Almenoff P. Determining population based mortality risk in the Department of Veterans Affairs. *Health Care Manag Sci.* 2012;15(2):121-137.
9. Tan A, Kuo YF, Goodwin JS. Predicting life expectancy for community-dwelling older adults from Medicare claims data. *Am J Epidemiol.* 2013;178(6):974-983.
10. Mathias JS, Agrawal A, Feinglass J, Cooper AJ, Baker DW, Choudhary A. Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data. *J Am Med Inform Assoc.* 2013;20(e1):e118-124.
11. Wang L, Porter B, Maynard C, et al. Predicting risk of hospitalization or death among patients with heart failure in the veterans health administration. *Am J Cardiol.* 2012;110(9):1342-1349.
12. Wang L, Porter B, Maynard C, et al. Predicting risk of hospitalization or death among patients receiving primary care in the Veterans Health Administration. *Med Care.* 2013;51(4):368-373.
13. Austin PC, Newman A, Kurdyak PA. Using the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a population-based cohort of adults with schizophrenia in Ontario, Canada. *Psychiatry Res.* 2012;196(1):32-37.
14. Austin PC, van Walraven C. The mortality risk score and the ADG score: two points-based scoring systems for the Johns Hopkins aggregated diagnosis groups to predict

- mortality in a general adult population cohort in Ontario, Canada. *Med Care*. 2011;49(10):940-947.
15. Austin PC, van Walraven C, Wodchis WP, Newman A, Anderson GM. Using the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a general adult population cohort in Ontario, Canada. *Med Care*. 2011;49(10):932-939.
 16. Quail JM, Lix LM, Osman BA, Teare GF. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC Health Serv Res*. 2011;11:146.
 17. Ogata K, Miyamoto T, Adachi H, et al. New computer model for prediction of individual 10-year mortality on the basis of conventional atherosclerotic risk factors. *Atherosclerosis*. 2013;227(1):159-164.
 18. Lee S, Lindquist K, Segal M, Covinsky K. Development and validation of a prognostic index for 4-year mortality in older adults. *JAMA*. 2006;295(7):801-808.
 19. Synnestvedt MB, Weiner MG. Evaluation of the use of electronic health data to classify four-year mortality risk for older adults undergoing screening colonoscopies. *Proc AMIA Symp*. 2011;2011:1361-1365.
 20. Stotter A, Reed MW, Gray LJ, Moore N, Robinson TG. Comprehensive Geriatric Assessment and predicted 3-year survival in treatment planning for frail patients with early breast cancer. *Br J Surg*. 2015;102(5):525-533.
 21. Tang V, Shi Y, Fung K, et al. Clinician factors associated with prostate-specific antigen screening in older veterans with limited life expectancy. *JAMA Intern Med*. 2016.
 22. Wyatt J, Altman D. Commentary: prognostic models: clinically useful or quickly forgotten? *BMJ*. 1995;311:1539-1541.
 23. Siontis G, Tzoulaki I, Ioannidis J. Predicting death: an empirical evaluation of predictive tools for mortality. *Arch Intern Med*. 2011;171(19):1721-1726.
 24. Moons K, Altman D, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.

APPENDIX A. SEARCH STRATEGY

MEDLINE (Ovid) through May 2016

- 1 (life adj expectancy).m_titl.
- 2 (survival or mortality or death).m_titl.
- 3 1 or 2
- 4 (calculat\$ or instrument\$ or index or indice\$ or model\$ or tool\$ or prognosis or risk or predict\$ or estimat\$).m_titl.
- 5 3 and 4
- 6 (valid\$ or calibrat\$ or compar\$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
- 7 5 and 6
- 8 limit 5 to (english language and humans and yr="2011 -Current" and ("middle-aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)"))
- 9 limit 7 to (english language and humans and yr="2011 -Current" and ("middle-aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)"))
- 10 limit 8 to (clinical trial or comparative study or controlled clinical trial or evaluation studies or meta analysis or observational study or pragmatic clinical trial or randomized controlled trial or systematic reviews or validation studies)
- 11 limit 9 to (clinical trial or comparative study or controlled clinical trial or evaluation studies or meta analysis or observational study or pragmatic clinical trial or randomized controlled trial or systematic reviews or validation studies)
- 12 8 NOT 9

APPENDIX B. PEER REVIEW COMMENTS/AUTHOR RESPONSES

Question	Reviewer Comment	Author Response
Are the objectives, scope, and methods for this review clearly described?	Yes	Thank you
	Yes	
	Yes	
Is there any indication of bias in our synthesis of the evidence?	No	Thank you
	No	
	No	
Are there any <u>published</u> or <u>unpublished</u> studies that we may have overlooked?	No	No response needed
	Yes - 1. Cruz et al JAMA research letter 2012 Predicting 10 year mortality in older adults. Granted, a research letter not a full research publication, but a high impact journal and an important update to the prior 4 year mortality index. 2. Lee et al PLOS one 2014 Individualizing life expectancy estimates using Gompertz... The conclusion of the report discuss lack of prognostic models that estimate life expectancy (time to death) rather than mortality risk (risk of death over a given time frame). This study estimates life expectancy, using the Lee index mentioned in 1 above. 3. Schonberg M JAGS 2011 External Validation of an index to predict 9 year...Granted, included in our review by Yourman, published in 2011 so I believe should have been included.	We have reviewed the suggested studies. Although the Cruz and Schoenborn prognostic models are potentially useful, these reports were not included in this review because the mortality prediction models are based on self-reported national survey data and would require similar patient questionnaires be administered by primary care clinics. We have included the Lee et al 2014 reference so readers can see how one could model survival curves to estimate life expectancy.
	No	No response needed
Additional suggestions or comments can be provided below. If applicable, please	p. 5 line 5 - should be "insubstantial" rather than "unsubstantial" p. 5 line 10 - "quasi-validation" should probably be defined in the table	We have made the suggested change on pg 5, line 5. 'Quasi' simply means the studies were

<p>indicate the page and line numbers from the draft report.</p>	<p>on p 1 (or avoided altogether as an imprecise term).</p>	<p>not true external validation studies, however they were in a sense validation studies. We have pointed that none of the studies reviewed for KQ2 were true validation studies, and dropped 'quasi' from the Table title. We did not feel a need to define quasi in the table on page 1.</p>
	<p>Major points:</p> <ol style="list-style-type: none"> 1. As senior author of the Yourman systematic review of prognostic indices for older adults, I'm pleased that this report has been commissioned and well executed. Many of the conclusions remain the same - people just are not studying the use/usefulness of prognostic indices in clinical practice. As a key member of the eprognosis team, a free website dedicated to making prognostic models available to clinicians, I can testify that prognostic models are being used in everyday practice. Clinicians come up to me all the time, from all over the country, and say "I use eprognosis." It seems to be more the specialists that use the models rather than the primary care clinicians, as a recent study by Nancy Shoenborn in JAMA Internal Medicine 2016 suggests. 2. The conclusions are based on the time limited update between 2011 and 2016. Yet the underlying question "are there clinically useful prognostic models for VA primary care patients" need not be restricted to this time frame. Certainly, risk may change with time. But have things changed that much between, say 2000, and 2016? My guess is no, the same factors that put persons at risk in 2000 are likely to put persons at risk in 2016. Age, gender, functional limitations. We haven't cured cancer. To really address the underlying question, the VA will need to consider the accuracy and pragmatic usefulness of prior studies not included in this report. 3. Why no mention of the CAN score? This is the giant elephant in the room that VA researchers are talking about. Apparently it's calculated for veterans and hard but possible to access in the medical record. What's the evidence? How useful is it? Omission is a major limitation as it will be the first thing on many VA clinician/researcher's minds when they think VA data and prognostic index. 4. Minor points -Explain what a "quasi-validation" study means. 	<p>The Schoenborn article is now mentioned in the Introduction of the Evidence Report to support an important statement that providers often don't share long-term prognostic assessments with their patients in part due to their uncertainty.</p> <p>We agree. The VA can refer to these previous reviews, thus we focused on more recent studies.</p> <p>We had not pointed out that the 2 Wang articles cited for KQ1 and KQ2 represent the VA Care Assessment Needs (CAN) model. We have now repeatedly made this connection in the text and table footnotes.</p>

	<p>-The two studies by Austin seem to really be one study. This question comes up immediately as all data in the tables about them are essentially identical. One shouldn't have to wait for a footnote to discover one study is simply developing a point score system using the same model. Not clear if this requires a separate citation and line in the table, as it's the same index. Might be better organized if by index rather than by publication.</p>	<p>'Quasi' is used to indicate the studies were not true external validation studies, however they were in a sense validation studies. We have pointed that none of the studies reviewed for KQ2 were true validation studies, and dropped 'quasi' from the Table title. We did not feel a need to define quasi in the table on page 1.</p>
	<p>Thank you for the opportunity to review this excellent ESP.</p> <p>The authors sought to review recent life expectancy calculators for use in primary care practice. In particular, they looked for new calculators published since 2011, external validations of those new calculators, and clinical applications of the calculators.</p> <p>Their research questions were specific, their methods were clear, and the document is well written. Their primary findings are also clear. There have been multiple calculators created, including one in VA. These calculators have had quite good testing characteristics generally, but have not been validated in external samples and have not had any study in clinical practice.</p> <p>The tables and figures were clear and will be a valuable summary for anyone doing this type of work.</p> <p>Their findings drive home what I think will be a central issue in health services and implementation science for the next decade or more. This is that Big Data will make risk prediction reasonably easy and accurate, but having prediction tools does not tell us what predictions are useful, how do we use them, and how do we integrate tools into the system to be useful to patients, providers, and healthcare leaders. These questions will likely be very important, but they've shown few people seem to be asking them.</p> <p>At the broadest level, I feel like the questions were very well formulated, but the underlying problems the questions were trying to solve were somewhat less so. Exactly how is this information going to be used inside or outside VA? If this question were answered more explicitly, some of their other choices may have been slightly different.</p>	<p>Thank you.</p> <p>As pointed out in the text, the current VA model known as the CAN score doesn't predict all-cause model and wasn't developed to estimate life expectancy or evaluated as such, but could be adapted</p>

	<p>For example, they found a good study by Wang (the senior author of which is an operational partner on this project) that used VA data. If VA wants to implement a life expectancy calculator, shouldn't they just use this tool? Are there things other tools did that a modified VA-based tool would want to do?</p> <p>For KQ2 and KQ3, they were unlikely to find validation studies since the studies included were already recent. Here they could have also used any study in Yourman's 2011 JAMA paper, which was one of the inspirations for this study. The team that wrote the JAMA paper has a website, eprognosis.org, that does some of the things that KQ3 wanted to address, though I don't know if they have validated it in any way. While adding older studies may have been reasonably out of scope for this project, it could have found some interesting information.</p> <p>Their search strategy was reasonable. They only used Medline, which can miss citations. Because this is not a meta-analysis, missing citations is not a terribly big deal, so I find this reasonable.</p> <p>While I know this isn't the primary point of an ESP, I do wish the authors expanded their interpretive Research Gaps section a little. I was curious what they felt about each of the different scores, how they differ, and what future developers or users could learn from them.</p> <p>A few small typos P8 "appropriate to to" P12 Moons2014 should be inside the period.</p>	<p>to do so.</p> <p>If an effort to guide the VA, we tried to focus on what the good performing models had in common, and what further evaluation would be needed to support their use.</p> <p>The P8 typo has been corrected and all references have been converted to superscript format.</p>
--	--	---

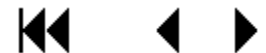
APPENDIX C. EVIDENCE TABLES

Table 1. Study Characteristics

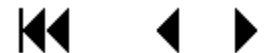
Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
Aggregated Diagnosis Groups (not VHA)					
Austin, 2011¹⁵ (General Adult Population) Canada, Ontario 2007-2008	Administrative healthcare databases in Ontario, Canada -Registered Persons Database (RPDB) -Discharge Abstract Database -Ontario Health Insurance Plan physician billing database -Ontario Mental Health Reporting System	Inclusion criteria: all persons in RPDB alive on their birthday in 2007 Exclusion criteria: age <20, age >100 Recruitment method: N/A	N=10,498,413 Age (years): 46 (median) Gender (% male): 49 Race: NR Predictors: Aggregated Diagnosis Groups [ADGs] with largest adjusted odds ratios; all OR>1.5 a) Psychosocial: recurrent or persistent, unstable (23.4%) b) Malignancy (5.8%) c) Chronic medical: unstable (17.1%) d) Time limited: major – primary infections (7.4%) e) Time limited: major (4.5%) f) Likely to recur: progressive (2.4%)	Definition: mortality within 365 days of index date (birthday in 2007) Measurement method: from RPDB (linked by encrypted health number) Duration of follow-up: 365 days Number of deaths: 85,007 (0.8%)	Predictor definition/measurement same for deceased/survivors: Yes Outcome definition/measurement same for deceased/survivors: Yes Outcome assessed independent of predictors (eg, blinded): Yes Incomplete follow-up or missing predictor data (% , handling of): NR Method of validation: Assessed over fitting by bootstrap methods

Author, Year Country, Region Study Dates	Source of Data^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
Austin, 2011¹⁴ (General Adult Population) Canada, Ontario 2007-2008	Administrative healthcare databases in Ontario, Canada -Registered Persons Database (RPDB) -Discharge Abstract Database -Ontario Health Insurance Plan physician billing database -Ontario Mental Health Reporting System	Inclusion criteria: all persons in RPDB alive on their birthday in 2007 Exclusion criteria: age <20, age >100 Recruitment method: N/A	N=10,498,413 Age (years): 46 (median) Gender (% male): 49 Race: NR Subgroup of 395,009 residing in rural areas Predictors: Aggregated Diagnosis Groups [ADGs] with largest adjusted odds ratios; all OR>1.5 a) Psychosocial: recurrent or persistent, unstable (23.4%) b) Malignancy (5.8%) c) Chronic medical: unstable (17.1%) d) Time limited: major – primary infections (7.4%) e) Time limited: major (4.5%) f) Likely to recur: progressive (2.4%)	Definition: mortality within 365 days of index date (birthday in 2007) Measurement method: from RPDB (linked by encrypted health number) Duration of follow-up: 365 days Number of deaths: 85,007 (0.8%) Rural subgroup: 4,464 (1.1%)	Predictor definition/measurement same for deceased/survivors: Yes Outcome definition/measurement same for deceased/survivors: Yes Outcome assessed independent of predictors (eg, blinded): Yes Incomplete follow-up or missing predictor data (% handling of): NR Method of validation: Assessed over fitting by bootstrap methods

Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
<p>Austin, 2012¹³ (Adult Population with Schizophrenia) Canada, Ontario 2007-2008</p>	<p>Administrative healthcare databases in Ontario, Canada -Registered Persons Database (RPDB) -Discharge Abstract Database -Ontario Health Insurance Plan physician billing database -Ontario Mental Health Reporting System</p>	<p>Inclusion criteria: all persons in RPDB alive on 1/1/ 2007 that had previous diagnosis of schizophrenia (295.x in ICD-9 or F20/F25 in ICD-10) Exclusion criteria: age <20, age >100 Recruitment method: N/A</p>	<p>N=94,466 Age (years): 47 (median) Gender (% male): 46 Race: NR Predictors: Aggregated Diagnosis Groups [ADGs] with largest adjusted odds ratios; all OR>1.5 a) Psychosocial: recurrent or persistent, unstable (71%) b) Time limited: major – primary infections (12%) c) Malignancy (4.9%) d) Likely to recur progressive (3.9%)</p>	<p>Definition: mortality within 365 days of index date (1/1/ 2007) Measurement method: from RPDB (linked by encrypted health number) Duration of follow-up: 365 days Number of deaths: 1915 (2.0%)</p>	<p>Predictor definition/measurement same for deceased/ survivors: Yes Outcome definition/measurement same for deceased/ survivors: Yes Outcome assessed independent of predictors (eg, blinded): Yes Incomplete follow-up or missing predictor data (% handling of): NR Method of validation: Application of previously developed model for general adult population</p>
<p>Charlson and/or Elixhauser Comorbidities</p>					



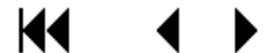
Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
<p>Gagne, 2011⁷ (Medicare enrollees, age ≥ 65)</p> <p>Pennsylvania, New Jersey, United States</p> <p>2004-2005</p>	<p>Inpatient & outpatient Medicare claims data & pharmacy databases in Pennsylvania and New Jersey for low-income Medicare enrollees who don't qualify for Medicaid</p> <p>Development cohort: Pharmacy Assistance Contract for Elderly (PACE) Pennsylvania</p> <p>Validation cohort: Pharmacy Assistance for the Aged and Disabled) PAAD) New Jersey</p>	<p>Inclusion criteria: Medicare enrollees with continuous drug coverage through PACE (development cohort) or PAAD (validation cohort) and at least one pharmacy claim during the 4 months before baseline year and survived the baseline year</p>	<p>NPA=120,679 Age 80 Gender (% female): 83 Race: NR</p> <p>NNJ=123,855 Age 79 Gender (% female): 77 Race: NR</p> <p>Predictors: Weighted comorbidity score calculated for 37 Romano/Charlson or vanWalraven/Elixhauser ICD-9 comorbidity classifications, age and gender; key predictors</p> <p>Heart failure:23% Dementia: 9.0% Renal failure: 6.9% Metastatic cancer 1.8% Weight loss: 1.5%</p>	<p>Definition: 1 year mortality (also had 30,90,180 day mortality)</p> <p>Measurement method: NR</p> <p>Duration of follow-up: up to 1 year</p> <p>Number of deaths: Development cohort N=10,769 (8.9%)</p> <p>Validation cohort N=9,230 (7.5%)</p>	<p>Predictor definition/measurement same for deceased/survivors: Yes</p> <p>Outcome definition/measurement same for deceased/survivors: NR</p> <p>Outcome assessed independent of predictors (eg, blinded): NR</p> <p>Incomplete follow-up or missing predictor data (% handling of): NR</p> <p>Method of validation: External sample</p>



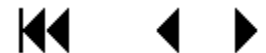
Author, Year Country, Region Study Dates	Source of Data^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
Quail, 2011¹⁶ Canada, Saskatchewan (provincial health plan) Fiscal years 2001-2002	Provincial Discharge Abstract Database (ICD-9 & 10 codes), Medical Services Database (ICD-9 codes), Population Registry, & Vital Statistics Registry Income from residence in census region	Inclusion criteria: residents age 20 and older with uninterrupted coverage in year (comorbidities) were assessed (FY2001) Exclusion criteria: federal employees, inmates, & First Nation people	General Population/Age > 65 years N= 662,423/137,700 Age: 48/75 Female 51%/57% Race: NR Charlson score 0.3/0.7 Elixhauser Heart failure: 2.0/8.1 Metastatic cancer:0.8/2.4 Renal failure: 0.6/1.8 Weight loss: 0.1/0.2 Pulmonary disease: 8.4/12.7	1-year mortality General Population/Age > 65 years 1.3%/ 5.1%	Predictor definition/measurement same for deceased/survivors: Yes Outcome definition/measurement same for deceased/survivors: Yes Outcome assessed independent of predictors (eg, blinded): Yes Incomplete follow-up or missing predictor data (% handling of): missing income imputed Method of validation: external sample & time

Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
<p>Tan 2013⁹ United States 1999-2009</p>	<p>5% random sample of Medicare enrollees in 2000</p> <p>Medicare enrollment files, carrier files, outpatient statistical analysis files (outpatient visits), Medicare provider analysis and review files (hospital stays)</p>	<p>Inclusion criteria: 66-90 years in 2000, full coverage in Medicare Parts A (hospital care) and B (physician and outpatient services) in 1999, and not in Medicare Advantage HMO coverage at any time in 1999</p>	<p>N=1,137,311 Medicare beneficiaries Women: 60% Age (SD) 76 (6.5) Men: 40% Age (SD) 75 (6.1)</p> <p>Prevalence of comorbidities with > 80% 10-year mortality rates in women/men (%) Heart failure: 8.2/8.8 Pulmonary circulation disease: 0.9/0.8 Metastatic cancer: 0.8/0.9 Renal failure: 1.3/2.0 Weight loss: 1.5/1.2 Neurological disorders: 2.9/3.3 Substance abuse: 0.2/0.5 Dementia: 2.6/1.7 Psychoses: 1.3/0.9</p>	<p>1-, 5-, 7-, 10-year mortality stratified by sex</p> <p>Measurement method: Medicare enrollment</p> <p>10-year mortality Women: 51% Men: 57%</p>	<p>Predictor definition/measurement same for deceased/survivors: Yes</p> <p>Outcome definition/measurement same for deceased/survivors: Yes</p> <p>Outcome assessed independent of predictors (eg, blinded): Yes</p> <p>Incomplete follow-up or missing predictor data (% handling of): NR</p> <p>Method of validation: random split sample</p>
<p>Electronic Medical Records</p>					

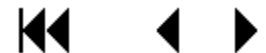
Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
<p>Mathias, 2013¹⁰ US, Chicago 2003-2008</p>	<p>EHRs (EpicCare) from a large, academic, multispecialty group practice and affiliated hospitals</p> <p>National Center for Health Statistics National Death Index (NDI) for 2003-2008</p>	<p>Inclusion criteria: outpatients, age 50 and older, at least 1 visit to the group practice during 2003</p> <p>Exclusion criteria: none reported</p> <p>Recruitment method: N/A</p>	<p>N=7,463 Age (years): 62 (mean) Gender (% male): 40 Race: White 51%; Black 24%; Hispanic 5%; Asian 3%; Other/Unknown 8%</p> <p>-Predictors: Count (mean[SD]) 2.5 (2.0) Present in >10% a) Hypertension 52% b) Any vascular disease 17% c) Diabetes 17% d) Any cancer 15% e) Tobacco use 11%</p> <p>-Vital signs: BP 131/79</p> <p>-Lab results: Albumin (g/dl) 3.7 (0.4) Creatinine (mg/dl) 1.1 (0.8)</p> <p>-Outpatient medications Digoxin 3% Loop diuretic 8%</p> <p>-Utilization PCP visits 1.2 (1.8) Hospitalizations in past year 0.3 (1.6); 1-2 years prior 0.2 (0.7)</p>	<p>Definition: death within 5 years of last outpatient encounter in 2003</p> <p>Duration of follow-up: 5 years</p> <p>Number of deaths: 838 (11.2%)</p>	<p>Predictor definition/measurement same for deceased/survivors: Yes</p> <p>Outcome definition/measurement same for deceased/survivors: Yes</p> <p>Outcome assessed independent of predictors (eg, blinded): Yes</p> <p>Incomplete follow-up or missing predictor data (% handling of): NR</p> <p>Method of validation: 10 fold cross validation</p>
<p>Data from Other Studies</p>					



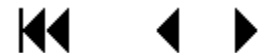
Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
<p>Ogata, 2013¹⁷ Japan; Tanushimaru, Uka 1999-2009</p>	<p>Baseline data collected for other study (one site was part of Seven Countries Study in Japan); reportedly similar to general population of Japan</p> <p>Survival or death from review of obituaries, medical records, death certificates, hospital charts, and interviews with primary care physicians, families, and other witnesses</p>	<p>Inclusion criteria: residents age >40 years, gave informed consent</p> <p>Exclusion criteria: missing data or lost to follow-up</p> <p>Recruitment method: "invited"</p>	<p>N=2,021 Age (years): 63 (mean) Gender (% male): 41 Race: NR (Japanese study)</p> <p>Predictors: 6 established cardiovascular risk factors Age (see above) Sex (see above) Systolic blood pressure: 134 mmHg HbA1c: 5.6% Total cholesterol: 200 mg/dl Current smoker 16%</p>	<p>Definition: survival or death within 10 years of baseline testing</p> <p>Duration of follow-up: 10 years</p> <p>Number of deaths: -Training sample: 204/1486 (14%) -Test sample 1: 49/365 (13%) -Test sample 2: 48/170 (28%)</p>	<p>Predictor definition/measurement same for deceased/survivors: Yes</p> <p>Outcome definition/measurement same for deceased/survivors: Yes</p> <p>Outcome assessed independent of predictors (eg, blinded): Unclear</p> <p>Incomplete follow-up or missing predictor data: 3.6% missing data/lost at site 1; 80% excluded at site 2 due to missing data/lost or < 10 years follow-up</p> <p>Method of validation: a) split sample (random) from first site and b) second site only served as validation site</p>
<p>Veterans Health Administration</p>					



Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
<p>Stefos, 2012⁸ US (VHA) Fiscal year 2007-2008</p>	<p>VA electronic administrative data files</p>	<p>Inclusion criteria: all patients who received care at a VA hospital in fiscal year 2008 that were assigned to a primary care provider</p> <p>Exclusion criteria: none reported</p> <p>Recruitment methods: N/A</p>	<p>N=4,774,000 Age: 62 Gender: 94% male Race: 55% white</p> <p>Key Predictors Categorizations of ICD-9 diagnosis codes from any type of encounter into Hierarchical Coexisting Conditions (HCC) using Diagnostic Cost Group (DCG) software (eg metastatic cancer, end-stage liver disease, respiratory arrest, coma of brain compression/anoxia) -Cancer: 17% -Chronic obstructive pulmonary disease: 10% -VA chronic disease registry: 4% -VA priority status (eg, catastrophically disabled (4%), low income or Medicaid (29%))</p>	<p>Definition: death from any cause within 12 months of last FY2008 VA clinical contact</p> <p>Measurement method: not clear</p> <p>Duration of follow-up: 1 year</p> <p>Number of deaths: 262,260 (5.5%)</p>	<p>Predictor definition/measurement same for deceased/survivors: Yes</p> <p>Outcome definition/measurement same for deceased/survivors: Yes</p> <p>Outcome assessed independent of predictors (eg, blinded): Yes (database)</p> <p>Incomplete follow-up or missing predictor data: NR</p> <p>Method of validation: Copas test for overfitting (repeated, split sample, cross validation design)</p>



Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
<p>Wang, 2013¹² (VHA primary care population for FY 2011)</p>	<p>Predictors from VHA electronic National Patient Care Database & Corporate Data Warehouse including demographics, diagnoses from inpatient & outpatient records, medications, vital signs, laboratory tests & healthcare utilization</p> <p>VHA Databases Death: VHA's vital status file</p>	<p>Inclusion criteria: All patients enrolled and assigned to a primary care provider within VHA on October 1, 2010 (the index date)</p> <p>Exclusion criteria: Patients with no recorded use of any health service during the prior year (5%); patients who were hospitalized, admitted to a hospital, or died on October 1, 2010</p>	<p>N=4,598,408 Age: mean 64 years range 18-110 Gender (male): 94% Race: NR</p> <p>Predictors (select): Charlson comorbidities & hierarchial condition categories Heart failure: 5.1% Renal failure: 6.3% Chronic pulmonary disease: 11% Metastatic cancer: 1.3% Substance abuse: 6.1%</p> <p>Heart rate > 100: 2.9% Respiration rate > 20: 2.4% Albumin < 3.5: 4.1%</p>	<p>Definition: death without hospitalization within 90 days or 1 year</p> <p>Measurement method: VHA Vital Status File</p> <p>Duration of follow-up: 1 year</p> <p>Number of deaths: 32,147 (0.7%) in 90 days 120,192 (2.8%) in 1 year</p>	<p>Predictor definition/measurement same for deceased/survivors: Yes</p> <p>Outcome definition/measurement same for deceased/survivors: Yes</p> <p>Outcome assessed independent of predictors (eg, blinded): Yes</p> <p>Incomplete follow-up or missing predictor data (% handling of): Yes</p> <p>Method of validation: randomly split sample</p>



Author, Year Country, Region Study Dates	Source of Data ^a	Participants	Participant Description	Mortality Outcome	Risk of Bias
Wang, 2012¹¹ (VHA heart failure population from 2009)	-Predictors from VHA electronic National Patient Care Database & Corporate Data Warehouse including demographics, diagnoses from inpatient & outpatient records, medications, vital signs, laboratory tests & healthcare utilization VHA Databases -Death: VHA's vital status file	Inclusion criteria: Heart failure diagnosis within VHA in year prior to June 1, 2009, the index date Exclusion criteria: NR	N=198,640 Age: mean 73 years Gender (male): 98% Race: NR Predictors (select): Charlson comorbidities & hierarchial condition categories Heart failure: 100% Renal failure: 25% Chronic pulmonary disease: 31% Metastatic cancer: 2.7% Dementia: 8.3% Respiration rate > 20: 6.7% Albumin < 2.5: 1.3% Heart rate > 85: 16%	Definition: death without hospitalization within 30 days or 1 year Measurement method: VHA Vital Status File Duration of follow-up: 1 year Deaths: 1,788 (0.9%) in 30 days, 14,103 (7.1%) in 1 year	Predictor definition/measurement same for deceased/survivors: Yes Outcome definition/measurement same for deceased/survivors: Yes Outcome assessed independent of predictors (eg, blinded): Yes Incomplete follow-up or missing predictor data (% handling of): Yes Method of validation: randomly split sample

^a eg, cohort, clinical trial participants, registry

^b Lee SJ et al. Development and validation of a prognostic index for 4-year mortality in older adults. JAMA. 2006;295(7):801-808.

BP = blood pressure; EHR = electronic health record; N/A = not applicable; NR = not reported; PCP = primary care provider; VA = Veterans Affairs; VHA = Veterans Health Administration

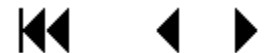
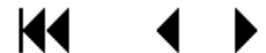


Table 2. Model Characteristics and Performance

Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
Aggregated Diagnosis Groups (not VHA)					
Austin 2011 ¹⁵ (General Adult Population)	Risk adjustment	<p><i>Demographic:</i> age, sex</p> <p><i>Patient History:</i> diagnoses associated with hospital admissions (ICD-10) and physician billing claims (ICD-9) from past 2 years matched with 32 Aggregated Diagnosis Groups (ADGs) (requires proprietary software license)</p> <p>Timing: Previous 2 years including index date</p>	<p>Logistic regression</p> <p>Backwards elimination- final model age, sex and 28 ADG's</p>	None pre-specified	<p>Predicted probabilities Range 0.00 to 0.90</p> <p>C-statistic <i>Validation cohort:</i> 0.92 Age ≥ 65: 0.81 Age < 65: 0.82</p> <p>Calibration Differences between observed and predicted mortality < 1% in all 100 centiles of predicted risk except top 3; biggest difference 3%</p> <p>Calibration plot Intercept 0.007 Slope 0.996</p>

Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
Austin 2011 ¹⁴ (General Adult Population)	Risk adjustment	<p><i>Demographic:</i> age, sex</p> <p><i>Patient History:</i> diagnoses associated with hospital admissions (ICD-10) and physician billing claims (ICD-9) from past 2 years matched with 32 ADGs (requires proprietary software license)</p> <p>Timing: Previous 2 years including index date</p>	<p>Mortality Risk Score (MRS) derived using regression coefficients of above final model including age and sex.</p> <p>Weighted ADG Score plus age and sex</p> <p>One- rather than 2-year look-back</p>	None pre-specified	<p>Predicted probabilities Range 0.00 to 0.90</p> <p>C-statistics <i>Validation cohort:</i> MRS: 0.92 for 1- and 2-year look-back period Rural subgroup: 0.90 ADG: 0.91 for 1- and 2-year look-back period</p> <p>Calibration Differences between observed and predicted mortality ~ 1% in 20 groups of predicted risk for both MRS and ADG</p> <p>Individual predicted probability of dying within 1 year increasingly under estimates observed mortality as predicted probabilities exceed ~0.2</p> <p>Calibration plots MRS: Intercept 0.007 Slope 0.996 Rural subgroup: Intercept 0.142 Slope 0.960 ADG: Intercept 0.006 Slope 0.996</p>



Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
<p>Austin, 2012¹³ (Adult population with schizophrenia) Canada, Ontario 2007-2008</p>	<p>Risk adjustment</p>	<p>Application of previously developed model for adult general population (See Austin 2011 a)</p>	<p>Application of previously developed model for adult general population (see Austin 2011 a) to subpopulation with schizophrenia</p>	<p>None pre-specified</p>	<p>Predicted probabilities NR for adult general population model C-statistic: 0.84 Calibration plot Intercept 0.356 Slope 0.0.805</p>
Charlson and/or Elixhauser Comorbidities					
<p>Gagne, 2011⁷ (Medicare enrollees, age ≥ 65) Pennsylvania, New Jersey, United States 2004-2005</p>	<p>Provide comorbidity score for risk adjustment</p>	<p><i>Demographic:</i> age, sex <i>Comorbidities:</i> ICD-9 diagnosis codes Recorded during a baseline year Jan 1, 2004 to Dec 31, 2004</p>	<p>Logistic regression to assign weights for 37 unique comorbidities; no variable selection Final weighted model 20 comorbidities with nonzero weights plus age, sex Comorbidity scores in validation cohort median: 1 interquartile range: 2 range: -2 to 18 zero: 27%</p>	<p>None pre-specified</p>	<p>Predicted probabilities; individual range NR; lower decile ~3% upper decile ~55% C-statistic <i>Validation cohort:</i> NJ 1 year 0.79 30 day 0.86 90 day 0.82 180 day 0.81 Calibration curve under predicted by ~ 3% in 5 % 6th deciles with 20 -25% mortality, over predicted by ~ 10% in 10th decile with ~ 45% mortality, otherwise observed and predicted mortality % within ~1%</p>



Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
Quail, 2011¹⁶ Canada, Saskatchewan (provincial health plan) Fiscal years 2001-2002	Population risk-adjustment	<i>Demographic:</i> patient: age, gender, income, & region <i>Comorbidities:</i> 17 in weighted Charlson score or 31 separate Elixhauser comorbidities based on diagnoses from baseline year (FY2001) from inpatient & outpatient records	Logistic regression model including age, sex, income, region plus Charlson comorbidity score or individual Elixhauser comorbidities fit to the data; no statistical variable selection	None pre-specified	Predicted probabilities: NR C-statistic general population/age > 65 years: Carlson model: 0.90/0.78 Elixhauser model: 0.91/0.80 Calibration: NR Prediction error (Brier score – mean squared difference between individual predicted probabilities and death =1 or survival=0) Carlson model: 0.01/0.04 Elixhauser model: 0.01/0.04

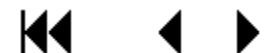
Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
<p>Tan 2013⁹ United States 1999-2009</p>	<p>Healthcare decision-making (expected to live long enough to benefit from a service such as cancer screening)</p>	<p><i>Demographic:</i> age, sex <i>Comorbidities:</i> 31 from Elixhauser, 17 from Charlson based on Quan coding algorithm of ICD-9 codes; appearing on 2 or more claims at least 30 days apart <i>Utilization:</i> number of hospital admissions, number of outpatient visits in previous 12 months (1999)</p>	<p>Variable selection: Series of logistic regression models with varying combinations of predictors; final model chosen based on best C-statistic, Akaike information criterion, and percent correctly classified Final model: Age + 31 individual Elixhauser comorbidities stratified by sex Sex-specific Cox proportional hazards models to generate K-M curves using median survival time as proxy of life expectancy</p>	<p>None prespecified; Predicted risk of death within 10 years categorized as <25%, 25-49%, 50-74%, and >75% by sex 5- and 10-year life expectancy often used for decisions about cancer screening</p>	<p>Predicted 10-year mortality: NR C-statistic Women/Men 10 years: 0.79/0.77 5 years: 0.78/0.76 1 year: 0.79/0.77 1-, 5-, and 10-year observed mortality fell within quartiles of predicted probabilities except 1-year observed mortality was less than predicted for 50% to 75% & ≥ 75% quartiles of risk for both women & men Positive predictive value (observed mortality by predicted life expectancy (LE using median survival time as proxy) was similar for women & men) LE < 10 years: 75% LE < 5 years: 69% LE < 1 year: 48%</p>
<p>Electronic Medical Records</p>					



Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
<p>Mathias, 2013¹⁰</p>	<p>Healthcare decision-making (eg, expected to live long enough to benefit from a service such as cancer screening)</p>	<p><i>Demographic:</i> 11 attributes; age, sex, marital status, race/ethnicity, socioeconomic status</p> <p><i>Comorbidities:</i> 117 attributes based on ICD-9 codes, current procedural terminology codes or substance use statuses; codes extracted from encounter diagnoses, past medical history, past surgical history, social history, and problem list; additional 26 attributes for counts of encounters related to frequent exacerbations of conditions or active diagnoses</p> <p><i>Vital signs:</i> 24 attributes of heart rate, SBP, DBP, pulse pressure</p> <p><i>Medications:</i> 664 medication attributes classified using VA codes; used medication list at index visit or medications ordered in year prior; added focus on some classes of medications</p> <p><i>Laboratory:</i> 120 laboratory attributes based on 24 tests</p> <p><i>Utilization:</i> 50 attributes; discharge status, hospital admissions, ED visits, home health referrals, provider visits</p> <p>Timing: year prior to index visit except home health referrals and provider visits included 1-2 years prior</p>	<p>Rotation forest ensembling technique with alternating decision tree</p> <p>Correlation Feature Selection (CFS) and manual review/reduction used to reduce number of attributes (eliminate low face validity, redundant, problematic reliability); information gain metric</p> <p>Final model: 24 predictors</p>	<p>Predicted risk of death within 5 years < 50% or ≥ 50%</p> <p>5-year life expectancy often used for decisions about cancer screening</p> <p>Predicted risk of 50% equivalent to median life expectancy of 5 years</p>	<p>Predicted Mortality Lowest risk decile 3.6% Highest risk decile 92.5%</p> <p>C-statistic 0.86 (0.85, 0.87)</p> <p>Sensitivity: 31% Specificity: 98% Positive predictive value: 63% Negative predictive value: 92% Correct predictions: 90%</p> <p>Calibration Difference between observed and predicted mortality <3% across all deciles of predicted risk</p>
<p>Data from Other Studies</p>					



Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c																						
<p>Ogata, 2013¹⁷ Japan; Tanushimaru, Uka 1999-2009</p>	<p>Predict if individual will live or die within 10 years using only cardiovascular risk factors</p>	<p><i>Demographic:</i> Age, gender, smoking status <i>Vital signs:</i> SBP <i>Laboratory:</i> HbA1c, total cholesterol</p>	<p>Supervised statistical pattern recognition with a minimum distance classifier to derive regression coefficients for 6 predictors preselected by authors Coefficient for total cholesterol is negative; opposite of Western cohorts</p>	<p>Live or die within 10 years</p>	<p>Predicted to die in 10 years in validation samples Site 1: 36/365 (9.9%) Site 2: 35/170 (20.6%) C-statistic for predicting survival and death Site 1: 0.83 Site 2: 0.85 Calibration: difference (in % dead) between observed and expected Site 1: 3.5% Site 2: 7.6%</p>																						
Veterans Health Administration																											
<p>Stefos, 2012⁸ US (VHA) Fiscal year 2007-2008</p>	<p>Estimate adjusted mortality statistics for VA hospital-based patient populations</p>	<p>Age, gender, VA priority status, marital status, race, insurance (<i>ie</i>, not insured by public or private insurance plan) Average driving time to 3 VA institutions (closest providing primary care, closest providing secondary/intermediate care, closest providing tertiary/specialty care) 139 Hierarchical Coexisting Conditions (HCCs) Membership in a VA Registry Program (special emphasis programs for specific chronic conditions) Timing of Assessment: VA administrative data from FY 2008</p>	<p>Hierarchical generalized linear mixed model with random effect for hospital population 2-stage estimation with insignificant ($P > .10$) covariates eliminated after first stage; final model included 14 demographic and 139 morbidity HCC measures</p>	<p>None pre-specified</p>	<p>Range of predictive probabilities 2.5% in lowest decile to 94% in highest decile C-statistic: 0.86 Calibration: Observed:Predicted (O/P) - number of deaths by risk decile</p> <table border="1" data-bbox="1638 1120 1953 1461"> <thead> <tr> <th>Predicted</th> <th>O/P ratio</th> </tr> </thead> <tbody> <tr><td>2.5%</td><td>0.94</td></tr> <tr><td>13%</td><td>1.12</td></tr> <tr><td>24%</td><td>1.12</td></tr> <tr><td>34%</td><td>1.06</td></tr> <tr><td>45%</td><td>0.98</td></tr> <tr><td>55%</td><td>0.95</td></tr> <tr><td>65%</td><td>0.91</td></tr> <tr><td>75%</td><td>0.87</td></tr> <tr><td>85%</td><td>0.87</td></tr> <tr><td>94%</td><td>0.86</td></tr> </tbody> </table>	Predicted	O/P ratio	2.5%	0.94	13%	1.12	24%	1.12	34%	1.06	45%	0.98	55%	0.95	65%	0.91	75%	0.87	85%	0.87	94%	0.86
Predicted	O/P ratio																										
2.5%	0.94																										
13%	1.12																										
24%	1.12																										
34%	1.06																										
45%	0.98																										
55%	0.95																										
65%	0.91																										
75%	0.87																										
85%	0.87																										
94%	0.86																										



Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
<p>Wang, 2013¹² (VHA primary care population from FY 2011)</p>	<p>Identify high-risk (of hospitalization or death without hospitalization) primary care patients who might benefit from care coordination and special management programs such as intensive case management, telehealth, home care, specialized clinics, and palliative care</p>	<p><i>Demographic:</i> age, sex, marital status, VHA enrollment priority</p> <p><i>Medical conditions:</i> Deyo-Charlson index (ICD-9) & hierarchical condition classification of diagnoses</p> <p><i>Vital signs:</i> blood pressure, heart rate, respiratory rate, BMI</p> <p><i>Prior year use of VHA health services:</i> indicators for categories of numbers & types of outpatient visits, ER visits, and hospitalizations over the past year and past month</p> <p><i>Medications dispensed:</i> number of refills, 31 types of medications</p> <p><i>Laboratory results:</i> Albumin, blood urea nitrogen, creatinine, potassium, white blood cell count</p>	<p>Multinomial logistic regression with 3 mutually exclusive categories: hospitalization, death without hospitalization, and neither event. Separate models for 90-day and 1-year endpoints.</p> <p>Backwards elimination followed by forward selection including select 2-way interactions.</p> <p>Final models contained up to 190 coefficients (numerous categorical variables had multiple coefficients)</p>	<p>None pre-specified</p>	<p>Predicted probabilities 90-day: lower decile <0.1% to upper decile ~4% 1-year: lower decile <0.1% to upper decile ~14% 1-year: age > 65 years: lower 5% <0.1%, upper 5% ~27%</p> <p><i>C-statistic</i> 90-day death: 0.86 1 year death: 0.85 1 year death >65 years: 0.80</p> <p><i>Calibration plots (Cox)</i> 90-day death: Intercept: -0.016 Slope: 0.999 1 year death: Intercept: 0.001 Slope: 1.002</p> <p>Small differences between observed and predicted in each decile of predicted probabilities; same if age > 65 years</p> <p>Observed deaths if in upper 5% of predicted probabilities of death without hospitalization 90-day: 6.2% 1-year: 19.4% 1-year age >65 years: 24.6%</p>



Author, Year	Intended Use	Predictors ^a Timing of Predictor Assessment ^b	Modelling Method Method for Selection of Predictors for Inclusion	Mortality Risk Groups	Predictive Performance ^c
Wang, 2012 ¹¹ (VHA heart failure population from 2009)	Identify high-risk (of hospitalization or death without hospitalization) patients with heart failure who might benefit from care coordination and special management programs such as intensive case management, telehealth, home care, specialized clinics, and palliative care	<p><i>Demographics:</i> age, sex, marital status, VHA enrollment priority</p> <p><i>Medical conditions:</i> Deyo-Charlson index (ICD-9) & hierarchical condition classification of diagnoses</p> <p><i>Vital signs:</i> blood pressure, heart rate, respiratory rate, BMI</p> <p><i>Prior year use of VHA health services:</i> indicators for categories of numbers & types of outpatient visits, ER visits, and hospitalizations over the past year and past month</p> <p><i>Medications dispensed:</i> number of refills, 31 types of medications</p> <p><i>Laboratory results:</i> Albumin, blood urea nitrogen, creatinine, potassium, white blood cell count</p>	<p>Multinomial logistic regression with 3 mutually exclusive categories: hospitalization, death without hospitalization, and neither event. Separate models for 30-day and 1-year endpoints.</p> <p>Backwards elimination followed by forward selection including select 2-way interactions such as age > 65 with medical conditions, medications & hospitalizations.</p> <p>Final models contained up to 190 coefficients (numerous categorical variables had multiple coefficients); same variables as other Wang report but regression coefficients and variable selection specific to this sample</p>	None pre-specified	<p>Predicted probabilities 30-day: lower decile <0.1% to upper decile ~4% 1-year: lower decile <0.8% to upper decile ~23%</p> <p><i>C-statistic</i> 30-day death: 0.80 1 year death: 0.76</p> <p><i>Calibration plots (Cox)</i> 30-day death: Intercept: -0.044 Slope: 1.000 1 year death: Intercept: -0.094 Slope: 0.96</p> <p>Differences between observed and predicted mortality % in each decile of predicted probabilities were not substantial</p> <p>Observed mortality rates if in upper 5% of predicted probabilities of death without hospitalization 30-day: 0.9 1-year: 0.34</p>

^a Definition and method of measurement

^b eg, at patient presentation, at event (retrospective)

^c Distribution of predicted probabilities, C-statistic, sensitivity/specificity for select cutpoints/risk groups, predicted/observed mortality in different risk groups, calibration slope, positive and negative predictive values

ADG = Aggregated Diagnosis Groups (Johns Hopkins); DBP = diastolic blood pressure; SBP = systolic blood pressure

