International Journal of General Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Evolving forecasting classifications and applications in health forecasting

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

This article was published in the following Dove Press journal:
International Journal of General Medicine
3 May 2012
Number of times this article has been viewed

Ireneous N Soyiri[1,2]
Daniel D Reidpath[1]

[1]Global Public Health, JCSMHS, MONASH University, Selangor, Malaysia; [2]School of Public Health, University of Ghana, Legon, Accra, Ghana

Abstract: Health forecasting forewarns the health community about future health situations and disease episodes so that health systems can better allocate resources and manage demand. The tools used for developing and measuring the accuracy and validity of health forecasts commonly are not defined although they are usually adapted forms of statistical procedures. This review identifies previous typologies used in classifying the forecasting methods commonly used in forecasting health conditions or situations. It then discusses the strengths and weaknesses of these methods and presents the choices available for measuring the accuracy of health-forecasting models, including a note on the discrepancies in the modes of validation.

Keywords: health forecast, health data, electronic health records, accuracy, cross validation, method, strengths and limitations

Introduction

Forecasting is the process of predicting future events based on foreknowledge acquired through a systematic process or intuition.[1,2] It requires data, information, and advanced knowledge. Forecasting has evolved over the years and now has wide applications in many fields, including economics and commerce,[1,3] sports,[4,5] the environment (including meteorology),[6,7] technology and politics,[8–10] and health.[11–14]

Health forecasting predicts health situations or disease episodes and forewarns about future events. It is also a form of preventive medicine or preventive care engaged in public health planning, and it is aimed at facilitating health care service provision in populations.[12,14–16] One of the least developed branches of forecasting, health forecasting is a useful tool for decision making in health services provision. Health forecasting has been commonly applied to emergency department visits, daily hospital attendance, and admissions.[17–20]

Various methods and approaches have been applied in forecasting events, but some outstanding issues are yet to be addressed. Even though a comprehensive classification of all forecasting approaches and methods would serve as a useful guide to forecasters searching for appropriate forecasting methods, there have been limited discussions and debates around this need.[21] Health forecasting studies have often adapted statistical techniques used by other well-established areas of forecasting, such as econometrics and finance. However, little has been said about the strengths and weaknesses of these techniques when they are applied to health forecasting.[22] Another important issue that has not been explicitly presented in the literature relates to approaches used to determine the accuracy and validity of health-forecasting models. The applications

Correspondence: Ireneous N Soyiri
Global Public Health, JCSMHS, MONASH University, 46150 Bandar Sunway, Selangor DE Malaysia
Tel +603 5514 6300 Extn 61569
Fax +603 5514 6323
Email soyiriin@yahoo.com

© 2012 Soyiri and Reidpath, publisher and licensee Dove Medical Press Ltd. This is an Open Access article which permits unrestricted noncommercial use, provided the original work is properly cited.

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

available for measuring and determining the validity and accuracy of health forecasts have not been compared or presented as guides for health forecasting or even used to stimulate discussions that can contribute to improve health forecasting. This paper therefore aims at presenting a brief overview of the evolution of forecasting classifications and methods. It discusses the strengths and weaknesses of various health-forecasting techniques and methods, and then presents the choices available for validating and measuring the accuracy of health-forecasting models. Because of the new approach it brings to medical and health sciences, health forecasting is important for practices in these fields. Advances in health-forecasting research will facilitate the decision-making processes that are associated with health-care planning and management.

In preparing this review, a search of the literature on health forecasting and statistical methods used in the analysis of health conditions was conducted in popular medical databases, such as PubMed (Medline) and Google Scholar. Additional literature searches were done through citation mapping of key papers. The selected papers and documents were synthesized and summarized according to the objectives of this paper.

An enumeration of forecasting typologies

Although several authors have made attempts to schematically classify the wide variety of forecasting techniques, many have not been adequately exclusive[a] or both concise and exhaustive[b] enough to meet the needs of forecasters across all fields.[21] A good classification system for forecasting methods can facilitate the process of choosing an appropriate method for forecasting, in addition to providing a better understanding and organization of the methods involved in designing a forecasting system. The enumeration of forecasting classifications presented below reveals the typologies and methods that have been involved. These classifications also justify the need for further research involving health-forecasting methodologies, since the latter have played a minimal role in shaping contemporary theory and methods in this area.

In 1971, Cetron and Ralph developed one of the earliest classifications of forecasting methods and approaches. It consisted of five categories, including intuitive methods, trend extrapolation, trend correlation, analogy, and dynamic predictive models.[23] Generally, intuitive methods in forecasting are based on individual opinion, whether structured or unstructured. Trend extrapolation is an approach that uses known existing trends, whereas trend correlation forecasts are based on the causal links between a dependent factor and another factor or factors. Cetron and Ralph also used the category, Analogy, to describe forecasting approaches that used similarity in patterns for forecasting. They also classified dynamic predictive models (also later known as structural models), which describe simulation procedures that involve high impact causal factors. Although Cetron and Ralph's classification is concise, it has been criticized for being neither exhaustive nor exclusive enough.[21]

Similarly, in 1972, in his classification of forecasting methods, Martino provided a five-category scheme consisting of the following: intuitive, consensus (ie, obtaining results from several experts), analogy, trend extrapolation, and structural models.[24] Although Martino's classification was concise and exclusive, it was not sufficiently exhaustive to meet the needs of forecasters. In 1978, another classification of forecasting methods by Bright considered as many as eight different categories, some of which were later thought contentious.[21,25] His classification included intuitive forecasting, trend extrapolation, dynamic modeling, morphological analysis, normative forecasting, monitoring, cross-impact analysis, and scenarios.[25] The key strength of Bright's classification was that it added an entirely new concept of scenarios, and could be viewed liberally as exhaustive. However, it was neither exclusive nor concise. Furthermore, as mentioned earlier, some categories, such as monitoring, have been challenged because they are inappropriate as forecasting methods or approaches.[21]

In 1985, Armstrong published his first "forecasting methodology tree," which was based on three assumptions or decisions. More recently, his *Methodology Tree for Forecasting* (2010) assumed that before arriving at an appropriate choice of analytical forecasting method, it is first necessary to decide on whether to use intuitive (judgmental) or objective (statistical) methods. Second, if the choice of approach is statistical, then a choice between causal and noncausal approaches is required. After a causal approach is chosen, the final decision is whether to select either linear or nonlinear (classification) statistical approaches.[26] Hence, Armstrong introduced five categories in his maiden classification: judgmental, bootstrapping, extrapolation, econometric, and segmentation. Armstrong's classification was concise and contributed new approaches to forecasting

[a]Exclusive means that anything belonging in one category should clearly not belong in another category.
[b]Exhaustive means that the classification system should cover every potential option.

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

(ie, naïve/causal continuum), as well as providing guidance on the selection of forecasting approaches, which previous classifications lacked. However, Armstrong's classification was neither exhaustive nor exclusive.

In 2001, Armstrong revised the classification of his forecasting methods and provided eleven categories of methods that could be derived from the *Methodology Tree* (including role playing, intentions, conjoint analysis, expert opinions, judgmental bootstrapping, analogies, extrapolation methods, rule-based forecasting, expert systems, econometric models and multivariate models).[1] His classification further illustrated the primary distinction between methods that rely on judgment[c] and those that estimate relationships using statistical[d] approaches or quantitative data.[1] The classification was however not concise because there were too many categories (eleven in all). It was also not exclusive because of the subclassification extrapolation, which has a much wider application in statistical forecasting.

In 2006, Gentry et al proposed an entirely new form of categorizing forecasting approaches and methods in the form of a grid. In *Forecasting Classification Grid*,[21] argue that two independent dimensions can determine forecasting approaches, which are on the continuums of Opinion and Empirical and Causal and Naïve. This classification helps to distinguish between opinions and ideas that can be empirically verified, and it is also simpler and more logical compared with earlier descriptions.[21] The *Grid* has just four classifications and is therefore a concise scheme. It also is exhaustive because it is designed to fit in as many forecasting methods as are available. Even though the authors believed that the classifications were exclusive, grey boundaries could exist between the continuums. The key challenge in this classification is that the grid appears to be an uncompleted template, so the exact relative positioning of a forecasting method could be the subject of debate.

In a 2010 revision of the *Methodology Tree for Forecasting*,[e] Armstrong and Green further extend the list of forecasting methods. The authors differentiated between structured and unstructured approaches related to judgmental forecasting and further classified the theory-based approaches of forecasting into the categories, linear and classification.[27] Armstrong and Green provided guidance on choosing suitable forecasting approaches and methods based on specific contexts. However, there are still flaws in their classification because some methods that had multiple applications were not explicitly illustrated in the *Methodology Tree*. For instance, both univariate and multivariate approaches can be applied in data mining or causal modeling.[28] Furthermore, noncausal (black box) approaches, such as those involving data mining and neural nets, are equally applicable to causal modeling and share similar methods and techniques (eg, regressions and segmentation).

All the classifications of forecasting approaches and methods discussed above have significantly contributed to the organization of forecasting. Even though most of these developments have taken place in nonhealth-related areas (eg, marketing, management and finance/econometrics), they have direct applications to health forecasting. For instance, health forecasting has used neural networks,[29–31] and many emergency department forecasts use one or more forms of regression analysis. It is therefore imperative that the lessons learned from previous forecasting topologies should inform any development of a typology of health-forecasting approaches and methods. Some related health-forecasting methods involved in the typologies listed above are exemplified in subsequent sections of this essay, which considers their strengths and limitations, their accuracy, and their validation procedures.

Health data and forecasting

Although data is vital in forecasting, what constitutes "health data" is poorly defined in the literature. Health data can be defined as: records of health conditions and situations that refer to individuals or populations and carry information about disease prevalence, incidence, diagnoses, treatments, prognosis, preventive strategies, and health systems. Moreover, these records are categorized by demographics and factors that directly affect health and are collected systematically or otherwise. For example, this definition could apply to hospital attendance or admission records that contain a variety of measures that are recorded in their respective units – age in complete years, for example.

In the practice of medicine, the diagnosis of disease is focused on determining the presence or absence of a condition so that the appropriate treatment can be given. However, the measures taken to facilitate this process are selected from a continuum. For example, diastolic blood pressure and pulse rate are taken to help determine whether a person is hypertensive or not (ie, more than or equal to 90 mmHg). Other factors that may have an effect on health status and whose levels

[c]Judgmental forecasting techniques include prediction markets, Delphi, structured analogies, game theory, decomposition, judgmental bootstrapping, and expert systems.

[d]Statistical forecasting approaches include causal models and segmentation.

[e]Also available at www.forecastingprinciples.com.

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

are measured to generate health data include environmental exposure (eg, weather and air quality). At the point of measure or use, health data could be classified as either continuous (ratio or interval scales) or categorical (ordinal, nominal, or dichotomous scales).[32] The definition and classification of health data determine how data are accumulated over time in addition to the method or methods of analyses that can be employed in analyzing this information.

An emerging form of health data – electronic health records (EHR) – refers to digital health data that is stored in secured repositories and shared only among authorized users.[33] Hayrinen and colleagues identified the following as components of electronic health records: daily charting, medication administration, physical assessment, admission nursing notes, nursing care plan, referral, present complaint (eg, symptoms), past medical history, life style, physical examination, diagnoses, tests, procedures, treatment, medication, discharge, history, diaries, problems, findings, and immunization.[34] These kinds of structured records have applications beyond health forecasting because they can be used to make predictions about the occurrence of future health events.

Strengths and limitations of health-forecasting techniques

Many reported studies on health forecasting adopted statistical techniques and methods, the theories of which are described in the standard literature. The choice of method depends on the purpose of forecasting and the nature of the data that are available. The strengths and limitations of these methods pertaining to health forecasting are discussed in the following paragraphs.

Linear regression methods are commonly used because they provide reasonably accurate results, are easy to interpret, and have wide applications in modeling trends and seasonality. However, like most regression methods, linear regression uses the method of ordinary least squares to derive estimates, which may wrongly assume that but for the dependent variable, the independent variables or regressors have no error.[35,36] Hence, to account for this problem in modeling, there is always a need to factor in an error component. Linear regressions also require large amounts of data on all variables for parameter estimations.[35]

Logistic regressions provide a means for analyzing binary or categorical dependent variables, but they are not useful for count data.[37] Logistic regressions can thus be applied to forecast the presence or absence of an event in a dichotomous (categorical) state. Poisson and negative binomial regressions are generally used for analyzing count data, and the latter is particularly suitable for analyzing count data that have a skewed distribution with a considerable number of zero entries.[38–42] For instance, Negative Binomial Models (NBMs) were used in previous work to investigate the determinants of asthma in the length of stay in hospitals, for which the dependent variable bore the aforementioned characteristics.[43,44] NBMs were also used to compare various statistical forecasting models for predicting the number of daily admissions of asthmatic patients in London (personal work yet to be published).

Moving average methods, which include autoregressive-integrated moving average (ARIMA), seasonal autoregressive-integrated moving average (SARIMA), and exponential smoothing (eg, Holt–Winters) are widely used forecasting approaches. They have the advantage of modeling trend and seasonal variations, as well as accommodating multivariable models.[45,46] The exponential smoothing methods used in health forecasting are effective with data that change over time.[22] However, the main challenge in using these complicated methods is that they require specialist knowledge and expertise.

Time series regressions generally have a much wider application and capabilities in forecasting than all the other nontime-series approaches mentioned here. Time series regressions provide easily interpretable outputs that can be more consistent than ordinary linear regressions.[45,46] The use of time series approaches in forecasting ideally requires sufficient data for not only the dependent variables, but also the matching independent variables.

Quantile regressions and fractional polynomials are rarely mentioned in the health-forecasting literature, but they provide a means for predicting and forecasting peculiar events.[47,48] For instance quantile regression models allow the modeling and forecasting of anticipated extreme events based on data distributions that are outside the normal range, which is more useful than linear regressions whose forecasts are based on the overall mean distribution. One major limitation of these approaches is that they deal with only the relevant or specific category of the data, and hence some information that could affect the accuracy and statistical power of the analysis is lost.[48,49]

Artificial neural network (ANN) is a black box modeling procedure known to provide more reliable results than the traditional causal approach.[50] ANN is capable of modeling complex and random systems by automatically controlling adjustments to the changes in time series based on the design of the experiment.[22,51] The major challenge of these models

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

is that they are difficult to interpret and, unlike the other approaches described earlier, very few statistical software packages are available.[52–54]

Measuring forecast accuracy and validation

Forecasting is generally aimed at predicting future events in order to inform and guide precautionary measures. It is an art in as much as it is a science, and therefore the degree of certainty of every forecast is imperative. A number of techniques and approaches are used to determine the accuracy or validation of a forecast. The main purpose of measuring the accuracy of a forecast model is to assist in choosing the best model.[55] This can be done in several ways: traditional forecasting accuracy measures; model discrimination approaches like receiver-operating characteristic (ROC) curves; and the use of model fit statistics, eg, R-square, Akaike information criterion (AIC), Schwarz information criteria (SIC) and Bayesian information criteria, which are discussed below.

Forecast accuracy

Forecast accuracy is a quantitative measure of the efficiency of the forecasting process, and it is performed by comparing the forecast to the actual situation. Forecast accuracy measures and parameters are usually supplied alongside the forecast as constituent elements to aid in decision making.

The accuracy of any forecast depends on objective features of the environment, such as the nature of the variable being forecast, or the length of the forecast horizon. Accuracy also depends on attributes of the forecast relating to the theories involved.[56] Hence, the choice of accuracy measures may depend on the method of forecasting. However, there has been considerable discussion about appropriate measures of forecasting accuracy, which have a wide range.[55,57–61]

Measures of forecasting accuracy have three main categories: (a) scale-dependent measures (ie, accuracy measures whose scale depends on the scale of the data); (b) percentage error measures (ie, independent measures that can compare forecast performance across different datasets); and (c) relative error measures (ie, scaled errors based on error measured from a reference standard forecast), including the relative measures of each type of error measure.[59] Examples of these measures are listed in Table 1.

Subsequent discussion focuses on selected scale-dependent and percentage error measures (Table 2), which are commonly used in health-forecasting studies. Scale-dependent error measures have been recommended for the comparison of different methods that are applied to the same set of data and scales.[59] For example, the root-mean-square error (RMSE) has traditionally been widely used for forecasting evaluation[62] and specifically for comparing models of the same series.[61,63] Even though some scale-dependent error measures, such as the MSE and RMSE, have been theoretically more relevant in statistical modeling, they have also been found more sensitive in detecting outliers than the mean absolute error (MAE) or median absolute error (MdAE).[59] Mean absolute scaled error (MASE) is also another scaled error approach recommended for comparing forecast accuracy across series on different scales.[59,64] According to Hyndman and Koehler, MASE provides the

Table 1 List of forecast accuracy measures

B. Scale-dependent measures
- I. Mean square error (MSE)
- II. Root mean squared error (RMSE)
- III. Mean absolute error (MAE)
- IV. Median absolute error (MdAE)

C. Percentage error measures
- I. Mean absolute percentage error (MAPE)
- II. Median absolute percentage error (MdAPE)
- III. Root mean square percentage error (RMSPE)
- IV. Root median square percentage error (RMdSPE)

D. Relative error measures[a]
- I. Mean relative absolute error (MRAE)
- II. Median relative absolute error (MdRAE)
- III. Geometric mean relative absolute error (GMRAE)

Notes: [a]The relative error measures are obtained by dividing each forecast error by the error obtained using a benchmark procedure, such as the grand mean (ie, a reference or benchmark average, which could be determined by taking the average of all averages of several subsamples). The accuracy measures of GMRAE and MDRAE, for instance, were presented by Armstrong and Collopy (1992)[61] and Fildes (1992).[86] Even though both reports recommend the use of forecast accuracy measures based on relative errors, they express these measures in different and complicated forms. Hyndman and Koehler (2006)[59] have however noted that these relative error methods could have some deficiencies that are associated with the difficulty of dealing with extremely small benchmark forecast error measures, resulting in the relative error measures having infinite variances.

Table 2 A comparison of scale-dependent error measures

Scale-dependent measures	Definition	Error spread	Error weights
Mean square error (MSE)	$\text{Mean}(O_t-F_t)^2$	Yes	Yes
Root mean squared error (RMSE)	$\sqrt{\text{MSE}}$	Yes	Yes
Mean absolute error (MAE)	$\text{Mean}\lvert(O_t-F_t)\rvert$	Yes	No
Median absolute error (MdAE)	$\text{Median}\lvert(O_t-F_t)\rvert$	–	–
Mean absolute scaled error (MASE)	$\text{Mean}\lvert(Q_t)\rvert$	Yes	Yes

Notes: Error spread refers to the ability of the measure to capture an error that is not localized and not widely distributed in the dataset. Error weights refers to the ability of the measure to differentiate the error at different points in history.
Abbreviations: t, at a time; O, observation; F, forecast; Q:A, scaled error independent of scale of data.[59]

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

most reliable approach because compared to others, it has a meaningful scale, is widely applicable, and is not subject to "degeneracy" problems.[59] Moreover, MASE is seen less sensitive to outliers and is more easily interpreted. It shows smaller variations, even with small samples, than other measures in the same category.[59,65]

Measures based on percentage errors are not dependent on the scales of data and hence can be used to compare forecast errors across different datasets. However, their results tend to be infinite or undefined if a given forecast result equates 0 at any given time or has an extremely skewed distribution when the forecast is close to 0.[59] A further challenge in this category of error measures, particularly for mean absolute percentage error (MAPE), is that they tend to over penalize positive errors compared to negative ones and thereby create an unbalanced symmetry in the measures.[66]

Model discrimination test (ROC curve)

The ROC curve is another measure of forecast error that is associated with discrimination and has been used in health related forecasting studies. ROC provides a means of measuring and comparing the accuracy of predictive models. It is a graphical plot of Sensitivity versus 1-Specificity in a binary classifier system, and it is constructed to assess the varying thresholds for discrimination of comparable statistical predictive models.[67–70] The accuracy of prediction is measured by comparing the true positives against false positives.[67,68] The ROC curve has very wide applications in many fields, and its use in forecasting has been described by many authors. It was for example used by Classen and Hokayem to compare various econometric models and to select a suitable model for forecasting "Childhood influences on youth obesity."[71]

Model fit statistics

Widely used statistical model fitness tests include R-square, adjusted R-square, AIC and SIC. These model parameters are defined and estimated as follows:

$$R^2 = 1 - \frac{RSS}{TSS} \tag{1}$$

$$\text{Adjusted } R^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{2}$$

$$AIC = e^{2k/n}\frac{RSS}{n} \tag{3}$$

$$SIC = n^{k/n}\frac{RSS}{n} \tag{4}$$

where RSS is the residual sum of squares; TSS is the total sum of squares; n is the sample size; k is the number of parameters in the fitted model.

The use of R^2 as a measure of fit or model variability in health-forecasting-related studies is very common in the literature.[20,72–75] A higher value of a model's R^2 could be interpreted as having a better fit, which tends to increase with the addition of every extra explanatory variable. However, using R^2 as a measure of fit can be unreliable in forecasting because the R^2 of a model can be high or equal to 1 and yet be consistently wrong.[57] Like the R^2, the adjusted R^2 also increases with every additional explanatory variable, but this test is more reliable because it tends to penalize the model for every additional explanatory variable as long as the new addition does not significantly reduce the RSS (Equation 2). The AIC is superior to the adjusted R^2 because it has a harsher penalty and is preferred in forecasting models as a measure of fit.[62,76,77] This technique is based on the maximum likelihood and the number of independently adjusted parameters within a predictive model.[76] Compared with AIC and R^2, SIC gives the best model diagnostic fit because it imposes the highest penalty on the model.[45] However, in forecasting, given the balance between the need for a predictive model that has a good fit and a high explanatory power, AIC is currently the most popular and recommended technique for model fit statistics,[78] and it is commonly used in model selection.[43,63,79–81]

Forecast validation

Forecast predictions are rarely perfect, so validation or cross-validation is an essential process that allows estimation of the extent to which a predictive model emulates the natural phenomenon that produces the data.[50,60,82] Validating a forecast requires appropriate techniques and reliable measures. In developing a health-forecasting predictive model, two types of validity can be examined: model validity and predictive validity. Both are important and can be used to generate a useful and reliable forecast. Model validity represents the extent to which the model fits the data that was used for the model development (ie, the fit of the model to the experimental sample). This type of validity test is also referred to in the literature as internal validity. The second type, predictive validity (also known as external validity), represents the extent to which the predicted forecast values fit the observed values (ie, the fit of the model

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

to the test sample).[61] Predictive validity is usually carried out through a process described as cross validation.

Cross validation is a statistical technique commonly used in forecasting for estimating the performance of a predictive model. It is usually carried out using a similar, but separate, sample of the data that was used in developing the forecast model. The health-forecasting literature does not provide standard procedures for conducting cross validation. Hence, the proportion of an evaluating sample (compared to the test or model development sample of data) that is suitable and sufficient for validating a health-forecasting model remains unclear. A scan of the literature revealed a wide range of arbitrary choices. As illustrated in Table 3, the relative proportion of a data sample used for cross validation of health-forecasting models could range from 1:1 to 12:1.

Table 3 Varying ratios of period of training to period of evaluation of health forecasting models

Author	Ratio of period of training: evaluation	Analytical techniques used in forecasting and study purpose
Hoot et al 2008[17]	1:1	ARIMA; to predict ED operation conditions within 8 hours
McCarthy 2008[18]	1:1	Poisson regression; to predict hourly ED presentations
Boyle 2011[19]	1:1/2:1/3:1/4:1	ARIMA, regression, ESM; to predict ED presentation and admission
Hoot et al 2007[51]	2:1	Logistic regression and ANN; to predict ED overcrowding
Wargon et al 2010[83]	3:1	Regression model; to predict daily ED presentation
Reis and Mandl, 2003[84]	4:1/5:1	ARIMA models; to predict daily pediatric ED presentation
Schweigler et al 2009[85]	7:1/14:1	SARIMA, hourly historical averages; to predict hourly ED bed occupancy
Jones et al 2008[22]	8:1	SARIMA, regression, ESM, and ANN; to predict daily presentation
Batal et al 2001[75]	9:1	Stepwise linear regression; to predict daily presentation
Champion et al 2007[20]	12:1	ARIMA, ESM; to predict aggregate monthly ED presentations

Abbreviations: ANN, Artificial Neural Networks; ARIMA, autoregressive-integrated moving average; SARIMA, seasonal autoregressive-integrated moving average; ESM, exponential smoothing; ED, emergency department.

For example, studies conducted by McCarthy et al[18] and Hoot et al[17] on forecasting emergency department visits used similar proportions of data for a cross validation of their output. However, other researchers have done this differently. To develop and to test their forecasts, Wargon et al[83] and Rotstein et al[74] used three quarters of their data as a training sample and the other one quarter as an evaluating sample.

Currently, there are no common scales for validating health forecasts based on a particular forecasting horizon, and the information available suggests that any appropriate cross validation strategy should be considered case by case. Thus, further research is necessary to help define and streamline the process of validating health-forecasting models.

Conclusion

The review identifies a number of knowledge gaps in health forecasting, which presents a challenge for further studies. These gaps include the following:

1. Typologies that classify health-forecasting approaches and methods;
2. A clear definition for health data, which nonetheless is an important ingredient for health forecasting;
3. Discussions on the strengths and limitations of statistical methods that are applicable to health forecasting, particularly for extreme health events;
4. A classification and ranking of various accuracy measures applicable to health forecasting; and
5. A clearly defined approach to cross validation of health-forecasting models.

The classifications of forecasting approaches have evolved over time. Several researchers have attempted to classify forecasting methods into typologies that are concise, exclusive, and exhaustive for all purposes. Lessons learned from these attempts will serve as useful guides in developing health-forecasting classification topologies and schemes, which are currently nonexistent. Few statistical methods have been identified to forecast extreme health events. Compared with percentage and relative error measures, scale-dependent error measures are easier and more frequently used in health forecasting. Because no common guidelines are available for cross validation in health forecasting, the current practice is quite irregular. Therefore, detailed studies are needed to help define standard classifications and applications for health forecasting.

Disclosure

The authors declare they have no competing or conflicts of interest in this work.

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

Authors' contributions

The paper was conceived and drafted by INS. Both INS and DDR participated in revising the draft. The final manuscript was approved by both authors.

Acknowledgments

We would like to thank the peer reviewers of this article for the useful comments and Ginarthini D/O Pachlai for editorial expertise.

References

1. Armstrong JS. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA: Kluwer Academic Publishers; 2001.
2. Lewis JB, McGrath RJ, Seidel LF. *Essentials of Applied Quantitative Methods for Health Services Managers*. Sudbury, MA: Jones and Bartlett Publishers, LLC; 2011.
3. Makridakis S, Taleb N. Decision making and planning under low levels of predictability. *Int J Forecast*. 2009;25(4):716–733.
4. Vaughan Williams L, Stekler HO. Sports forecasting. *Int J Forecast*. 2010;26(3):445–447.
5. Vaughan Williams L. *Information Efficiency in Financial and Betting Markets*. Cambridge, MA: Cambridge University Press; 2005.
6. McMichael AJ, Campbell-Lendrum DH, Corvalán CF, et al, editors. *Climate Change and Human Health: Risks and Responses*. Geneva, Switzerland: World Health Organization; 2003.
7. World Meteorological Organization. *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation*. WCDMP-No 72. Geneva, Switzerland: World Meteorological Organization; 2009.
8. Manton KG, Soldo BJ, Vierck E. Death and taxes: a contrary view. *Popul Today*. 1984;12(11):2, 8–9.
9. Rogge JR. Population trends and the status of population policy in Africa. *J Geog*. 1982;81(5):164–174.
10. Fildes R. The forecasting journals and their contribution to forecasting research: citation analysis and expert opinion. *Int J Forecast*. 2006;22(3):415–432.
11. Marno P, Chalder M, Laing-Morton T, Levy M, Sachon P, Halpin D. Can a health forecasting service offer COPD patients a novel way to manage their condition? *J Health Serv Res Policy*. 2010;15(3):150–155.
12. Met-Office. The Met Office health forecasting services. 2009. Available from: http://www.metoffice.gov.uk/health/. Accessed January 1, 2009.
13. World Health Organization. *Using Climate to Predict Infectious Disease Epidemics*. Geneva, Switzerland: World Health Organization; 2005.
14. Rogers DP, Shapiro MA, Brunet G, et al. Health and climate – opportunities. *Procedia Environmental Sciences*. 2010;1:37–54.
15. Van Meijgaard J, Fielding J, Shimkhada R, Eslami E, Cole B. *Forecasting Health Outcomes Related to Air Quality for Placer County. Issue Brief*. Los Angeles, CA: Health Forecasting, UCLA School of Public Health; 2010.
16. Sekhri N, Chisholm R, Longhi A, et al. Principles for forecasting demand for global health products. Global Health Forecasting Working Group Background Paper. 2006. Available from: http://www.cgdev.org/doc/DemandForecasting/Principles.pdf. Accessed February 2, 2009.
17. Hoot NR, LeBlanc LJ, Jones I, et al. Forecasting emergency department crowding: a discrete event simulation. *Ann Emerg Med*. 2008;52(2):116–125.
18. McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med*. 2008;15(4):337–346.
19. Boyle J, Jessup M, Crilly J, et al. Predicting emergency department admissions. *Emerg Med J*. Jun 24, 2011; [Epub ahead of print.]
20. Champion R, Kinsman LD, Lee GA, et al. Forecasting emergency department presentations. *Aust Health Rev*. 2007;31(1):83–90.
21. Gentry L, Calantone RJ, Cui SA. The forecasting classification grid: a typology for method selection. *J Global Bus Manage*. 2006; 2(1):48–60.
22. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15(2):159–170.
23. Cetron MJ, Ralph CA. *Industrial Applications of Technological Forecasting: Its utilization in R&D management*. New York, NY: Wiley-Interscience; 1971.
24. Martino JP. *Forecasting the Progress of Technology*. 2nd ed. New York, NY: Gordon and Breach Science Publishers; 1972.
25. Bright JR. *Practical Technology Forecasting: Concepts and Exercises*. Austin, TX: Industrial Management Center, Inc; 1978.
26. Martino JP. *An Introduction to Technological Forecasting*. London, UK: Gordon and Breach; 1972.
27. Armstrong JS, Green KC. JSA-KCG methodology tree for forecasting – forecasting principles. *Forecasting Principles*. 2010. Available from: http://www.forecastingprinciples.com/. Accessed March 19, 2011.
28. Fildes R, Nikolopoulos K, Crone S, Syntetos A. Forecasting and operational research: a review. *J Oper Res Soc*. 2008;59(9):1150–1172.
29. Blaisdell CJ, Weiss SR, Kimes DS, et al. Using seasonal variations in asthma hospitalizations in children to predict hospitalization frequency. *J Asthma*. 2002;39(7):567–575.
30. Kassomenos P, Papaloukas C, Petrakis M, Karakitsios S. Assessment and prediction of short term hospital admissions: the case of Athens, Greece. *Atmos Environ*. 2008;42(30):7078–7086.
31. Moustris KP, Douros K, Nastos PT, et al. Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. *Int J Environ Health Res*. 2012;22(2):93–104.
32. Bland JM. *An Introduction to Medical Statistics*. 3rd ed. Oxford, UK: Oxford University Press; 2006.
33. ISO/TR 20514. Health Informatics – Electronic Health Record – Definition, Scope, and Context: International Organization for Standards (ISO). 2005. Available from: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39525. Accessed March 10, 2012.
34. Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform*. 2008;77(5):291–304.
35. Cook RD, Weisberg S. Criticism and influence analysis in regression. *Sociol Methodol*. 1982;13:313–361.
36. Armstrong J. *Illusions in Regression Analysis*. Available from: http://ssrn.com/abstract=1969740. Accessed on December 8, 2011.
37. Collett D. *Modelling Binary Data*. 2nd ed. London, UK: Chapman and Hall/CRC Press; 2003.
38. Zorn C. Evaluating zero-inflated and hurdle Poisson specifications. *The Society for Political Methodology ad indicia spectate* 1998; Working Papers. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted. Accessed November 1, 2008.
39. Long JS. Predicted probabilities for count models. *Stata J*. 2001;1(1):51–57.
40. Anon. A brief overview of count data commands in Stata. 2008. Available from: http://www.sts.uzh.ch/past/hs08/count/stata_glossar.pdf. Accessed November 1, 2008.
41. Alt JE, King G, Signorino C. Aggregation among binary, count, and duration models: estimating the same quantities from different levels of data. *Political Analysis*. 1999;9:21–44.
42. Mitchell SM, Moore WH. A new look at cold war presidents' use of force: aggregation bias, truncation, and temporal dynamic issues. *The Society for Political Methodology Ad Indicia Spectate* 2000; Working Papers. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted. Accessed November 1, 2008.

International Journal of General Medicine downloaded from https://www.dovepress.com/ by 3.227.235.183 on 12-Apr-2021
For personal use only.

43. Soyiri I, Reidpath D, Sarran C. Determinants of asthma length of stay in London hospitals: individual versus area effects. *Emerg Health Threats J*. 2011;4(0):143–143.
44. Soyiri IN, Reidpath DD, Sarran C. Asthma length of stay in hospitals in London 2001–2006: demographic, diagnostic and temporal factors. *PLoS One*. 2011;6(11):e27184.
45. Shumway RH, Stoffer DS. *Time Series Analysis and its Applications with R Examples*. 2nd ed. New York, NY: Springer; 2006.
46. Chatfield C. *The Analysis of Time Series: An Introduction*. 6th ed. London, UK: CRC Press; 2004.
47. Koenker R. *Quantile Regression*. New York, NY: Cambridge Press; 2005.
48. Williams JS. Assessing the suitability of fractional polynomial methods in health services research: a perspective on the categorization epidemic. *J Health Serv Res Policy*. 2011;16(3):147–152.
49. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat*. 1994;43(3):429–467.
50. Breiman L. Statistical modeling: the two cultures. *Stat Sci*. 2001;16(3):199–215.
51. Hoot NR, Zhou C, Jones I, Aronsky D. Measuring and forecasting emergency department crowding in real time. *Ann Emerg Med*. 2007;49(6):747–755.
52. Spencer SJ, Alun T, Evans RS, Shari JW, Peter JH, Gregory LS. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15(2):159–170.
53. Lu WZ, Wang WJ. Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere*. 2005;59(5):693–701.
54. McKendry IG. Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting. *J Air Waste Manag Assoc*. 2002;52(9):1096–1101.
55. Mahmoud E. Accuracy in forecasting: a survey. *J Forecast*. 1984; 3(2):139–159.
56. Bathcelor R, Dua P. Forecaster ideology, forecasting technique, and the accuracy of economic forecasts. *Int J Forecast*. 1990;6(1):3–10.
57. Reidpath DD, Diamond MR, Hartel G, Glasziou P. Improving interpretability: as an alternative to R2 as a measure of effect size. *Stat Med*. 2000;19(10):1295–1302.
58. Lim C, McAleer M. Time series forecasts of international travel demand for Australia. *Tourism Management*. 2002;23(4):389–396.
59. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast*. 2006;22(4):679–688.
60. Willmott CJ. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*. 1982;63(11):1309–1313.
61. Armstrong JS, Collopy F. Error measures for generalizing about forecasting methods: empirical comparisons. *Int J Forecast*. 1992;8(1):69–80.
62. Barnett AG, Dobson AJ. *Analyzing Seasonal Health Data*. London, UK: Springer; 2009.
63. Hu W, Tong S, Mengersen K, Connell D. Weather variability and the incidence of cryptosporidiosis: comparison of time series poisson regression and SARIMA models. *Ann Epidemiol*. 2007;17(9):679–688.
64. Smyth C, Mori J, Jiang C. Model ensembles for prediction of Wenchuan aftershock activity. *Bulletin of the Seismological Society of America*. 2010;100(5B):2532.
65. Koehler AB, Snyder RD, Keith Ord J, Beaumont A. A study of outliers in the exponential smoothing approach to forecasting. *Int J Forecast*. 2012;28(2):477–484.
66. Makridakis S. Accuracy measures: theoretical and practical concerns. *Int J Forecast*. 1993;9(4):527–529.
67. Pepe M, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. *Stata J*. 2009;9(1):1.
68. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861–874.
69. Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. *Stat Med*. 2006;25(3):481–493.
70. Gorr WL. Forecast accuracy measures for exception reporting using receiver operating characteristic curves. *Int J Forecast*. 2009;25(1):48–61.
71. Classen T, Hokayem C. Childhood influences on youth obesity. *Econ Hum Biol*. 2005;3(2):165–187.
72. Tandberg D, Qualls C. Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med*. 1994;23(2):299–306.
73. Holleman DR, Bowling RL, Gathy A. Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. *J Gen Intern Med*. 1996;11(4):237–239.
74. Rotstein Z, Wilf-Miron R, Lavi B, Shahar A, Gabbay U, Noy S. The dynamics of patient visits to a public hospital ED: a statistical model. *Am J Emerg Med*. 1997;15(6):596–598.
75. Batal H, Tench J, McMillan S, Adams J, Mehler PS. Predicting patient visits to an urgent care clinic using calendar variables. *Acad Emerg Med*. 2001;8(1):48–53.
76. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19(6):716–723.
77. Bisgaard S, Kulahci M. Quality Quandaries*: Time series model selection and parsimony. *Qual Eng*. 2009;21(3):341–353.
78. Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecast*. 2002;18(3):439–454.
79. Beyerlein A, Fahrmeir L, Mansmann U, Toschke AM. Alternative regression models to assess increase in childhood BMI. *BMC Med Res Methodol*. 2008;8:59.
80. Lindsey JK, Jones B. Choosing among generalized linear models applied to medical data. *Stat Med*. 1998;17(1):59–68.
81. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. 2011;305(15):1553–1559.
82. Fildes R, Kourentzes N. Validation and forecasting accuracy in models of climate change. *Int J Forecast*. 2011;27(4):968–995.
83. Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med*. 2010;17(9):970–978.
84. Reis B, Mandl K. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak*. 2003;3(1):2.
85. Schweigler LM, Desmond JS, McCarthy ML, Bukowski KJ, Ionides EL, Younger JG. Forecasting models of emergency department crowding. *Acad Emerg Med*. 2009;16(4):301–308.
86. Fildes R. The evaluation of extrapolative forecasting methods. *Int J Forecast*. 1992;8(1):81–98.

International Journal of General Medicine

Dovepress

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. A key focus is the elucidation of disease processes and management protocols resulting in improved outcomes for the patient.The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: http://www.dovepress.com/international-journal-of-general-medicine-journal