ABSTRACT Breast cancer is a public health challenge in the Islamic Republic of Iran. Relatively few studies have dealt with large dataset and advanced statistical methods. Therefore, we aimed to investigate the role of prognostic factors on breast cancer survival using Additive Empirical Bayesian model with large data set. Breast cancer data set included 1574 women diagnosed with breast cancer from 2002 to 2012 that registered from Cancer Registry in Fars Province, Islamic Republic of Iran. Overall survival rates at 2, 3, 5 and 10 years were 0.98, 0.94, 0.87 and 0.76, respectively. Five years survival at stages 1, 2 and 3 were 0.94, 0.92 and 0.74, respectively. The younger patients with characteristics such as zero involved nodes, negative progesterone receptor, free skin and good prognostic level had a higher survival chance than others. The 5-year survival probability by stage in Fars Province was nearly the
same as that reported by the American Cancer Society. The Nottingham prognostic index (NPI) related to nodal status, tumour size and nuclear grade was the main indicator of breast cancer mortality.

Facteurs pronostiques du taux de survie du cancer du sein dans le sud de la République islamique d’Iran : approche bayésienne de type additif

RÉSUMÉ Le cancer du sein est un problème de santé publique en République islamique d’Iran. Relativement peu d’études ont traité la question en recourant à de larges jeux de données et à des méthodes statistiques avancées. Ainsi, nous avons voulu étudier le rôle des facteurs pronostiques de la survie du cancer du sein à l’aide d’une approche bayésienne de type additif et au moyen d’un large jeu de données. Le jeu de données sur le cancer du sein incluait 1574 femmes diagnostiquées avec un cancer du sein entre 2002 et 2012 inscrites au Registre du Cancer de la Province de Fars, en République islamique d’Iran. Au total, les taux de survie à deux, trois, cinq et 10 ans étaient de 0,98, 0,94, 0,87 et 0,76 respectivement. Les taux de survie à cinq ans aux stades 1, 2 et 3 étaient de 0,94, 0,92 et 0,74 respectivement. Les patients plus jeunes ayant des caractéristiques telles que l’absence d’implication ganglionnaire, la présence de récepteur négatif de la progestérone, l’absence d’atteinte cutanée, et un bon niveau pronostique avaient des chances de survie plus élevées que les autres. La probabilité de survie à cinq ans selon le stade dans la province de Fars était presque la même que celle rapportée par l’American Cancer Society. L’Index pronostique de Nottingham lié au statut ganglionnaire, à la taille de la tumeur et au grade nucléaire était l’indicateur principal de mortalité liée au cancer du sein.

Introduction

Breast cancer is the most prevalent cancer in the world with 1.7 million new cases and 521,900 deaths per year, making it the leading cause of cancer mortality among women. Around 25% of all cancer cases and 15% of all cancer deaths are associated with female breast cancer (1). It is the most frequent type of cancer among Iranian women, accounting for 24.6% of all cancers.
The mean age of females who have breast cancer is 49.6 years (2).

The incidence of breast cancer in the Islamic Republic of Iran increased from a crude rate of 22.0 per 100,000 women in 2005 to 22.6 per 100,000 women in 2015. In the last decade, the prevalence in this same population was 120 per 100,000 (2,3). The number of cases of breast cancer in Iranian women is smaller than that in neighbouring low- and middle-income countries (2). The number of deaths from breast cancer increased from 1.97 per 100,000 in 2006 to 2.45 per 100,000 in 2010. Previous research has presented the overall 5-year survival rate to be 71.0%; this rate varies according to the stage of the cancer (4).

Prognostic factors are imperative in the estimation of outcomes and the determination of optimal forms of treatment. In this context, prognostic factors should be differentiated from predictive factors. A prognostic factor is defined as a measurable variable that correlates with the natural history of the disease. In contrast, a predictive factor is one that is related to a given therapy in terms of response (5). Recent studies have demonstrated that influential prognostic factors that determine the survival time of breast cancer patients are: stage of disease, nodal status, tumour size, tumour grade, type of tumour, lymphatic and vascular invasion, hormone receptor status, Her2 (human epidermal growth factor receptor 2) status, Nottingham Prognostic Index (NPI) and age (5–11). The NPI is a numerical value that is computed by adding the values of tumour size (multiplied by a coefficient of 0.2), nuclear grade (1–3) and nodal status (1–3). The original NPI used a system which was classified into 3 categories identifying good, moderate and poor prognostic groups according to cut-off points pegged between the values 5.4, respectively (12). Higher scores have been associated with lower long-term survival rates.

Recently, data-based statistical models have been applied in the analysis of prognostic factors for the survival of patients with breast cancer in southern Islamic Republic of Iran (9,11). Survival analysis is a family of statistical models for data analysis in which the outcome variable is the time it takes for an event to occur. The main property that differentiates this analysis from other areas in statistics is censoring (13). The most popular survival analysis method is the Cox proportional hazard model. This states that covariates are multiplicatively related to the hazard (13). Sometimes evaluation of the effect of covariates on absolute change or survival function is a subject of interest for researchers. This willingness and violation of the proportionality hazard assumption of the Cox model guides the researcher to alternative models for which the additive hazard regression model may be more appropriate. Bayesian survival analysis consists of data and prior information; it generates conclusions based on the synthesis of new information from observed data and historical information. Empirical Bayesian methods are procedures for statistical inference in which prior distribution is estimated from the data while in the full Bayesian method, the prior distribution is fixed before any data are observed. The empirical Bayesian method has been shown to be an effective data analysis tool in recent years (14).
Although numerous studies have been conducted on the prognostic factors of breast cancer throughout the world, survival studies for a large number of patients in the Islamic Republic of Iran are rare, particularly studies conducted employing advanced statistical analysis. In the present study, advanced statistical analysis was used for a large number of Iranian breast cancer patients.

The objectives of the study were to estimate the survival probability at 2, 3, 5 and 10 years, and determine the effects of nodal status, nuclear grade, age, hormonal receptors, nipple involvement, skin involvement, lymphatic and vascular invasion, tumour size and NPI on breast cancer mortality using the additive empirical Bayesian hazard model. The outcomes of this study could help health organizations in projecting interventions for breast cancer.

**Methods**

**Study setting**

The Islamic Republic of Iran has a population of over 80 million, and most of the studies on breast cancer have been performed in the capital city, Tehran, which has a population of around 8.3 million, and reaching 14 million in the wider metropolitan area. However, the focus of most of those studies was not on the natural course and prognostic factors for breast cancer. In the southern part of the country, which has a population of almost 4 million, few advanced studies have been carried out on breast cancer. Our study was conducted in Shiraz in Fars province. The Cancer Registry Centre of Nemazi Hospital in Shiraz, a population-based cancer registry founded in 1971, is responsible for recording breast cancer in people living in the southern part of the country.

**Data source and subjects**

A cohort study on breast cancer patients was conducted in the Cancer Registry Centre of Nemazi Hospital, between January 2002 and December 2012. Inclusion criteria included all patients who met the following conditions: diagnosis with primary breast cancer within the period of 2002–2012, and having no other kind of cancer. A total of 1574 cases were selected for entry into the study. Medical records in the registry provided information on age, tumour size, skin involvement, estrogenic receptor, progesterone receptor, nipple involvement, node total, nuclear grade, lymphatic and vascular invasion (LV involvement), stage of disease, marital status and education. The type of therapy employed for all patients was chemotherapy. These patients were followed up in their community until December 2012.

The survival time of each patient was defined as the difference between the date of primary
diagnosis of breast cancer and the date of death from breast cancer only, or the closing date of follow-up. Death due to other reasons was considered a censored result in the study. The censoring mechanism considered in this study was generalized type I censoring, all patients were included until the end of study, none were lost to follow-up and each individual had a specific fixed censoring time.

**Prognostic factors**

We examined the effects of age, tumour size, skin involvement, estrogenic receptor, progesterone receptor, nipple involvement, nodal status, nuclear grade, lymphatic and vascular invasion, stage of disease, marital status, education and NPI on the survival time. Age at prognosis was used as a continuous variable for the sample. Marital status was recorded as married or single (women who had never been married before the time of diagnosis). Education status was categorized into 4 groups: illiterate, primary school, high school and university. The TNM staging system was used to classify the patients into stage at diagnosis: stage I, stage II (including stage IIA and IIB), and stage III (including stage IIIA, IIIB and IIIC) (there were no patients at stage IV). Nodal status was classified into 3 categories: 0, 1–3, and > 3 nodes involved. Tumour size was classified into 3 groups: 5 cm. The nuclear grade was stratified into 3 stages: well differentiated, poorly differentiated and undifferentiated. Nipple, skin and LV involvement were categorized into 2 groups, involved and free. Oestrogen and progesterone receptors were positive or negative. The NPI was divided into 3 categories: good, moderate and poor prognostic groups, with cut-off values ≤ 3.4, 3.5–5.4 and > 5.4 (12).

**Statistical analysis**

The overall survival rate of patients with breast cancer at 2, 3, 5 and 10 years was computed using the Kaplan–Meier method. In the process of analysis to determine the influential prognostic factors, the log-rank test and univariate Cox regression model were used to check the independence between survival time and each prognostic factor. Covariates with P-value 0.05).

Estimates of the regression coefficients for prognostic factors indicated that the risk of death in patients with more than 3 involved nodes increased compared with those without involved nodes. Survival decreased with older age at diagnosis. Risk of death for patients with negative progesterone receptor and skin involvement increased; and NPI had the strongest effect on the survival rate of patients (Table 2).

In the model checking process, the additive empirical Bayesian hazard model and Cox regression model are compared using goodness of fit criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) when both models are equivalent in
terms of number of covariates. The values for AIC in the present model and the Cox model were 1626.08 and 2547.28, respectively; the values for BIC were 1727.50 and 2616.98, respectively, indicating that the present model performed better than the Cox proportional hazard model. Comparison of the model-based survival functions against the observed survival functions also confirmed the model used in this study (Figure 4).

Discussion

The results of the current study showed similarities in the epidemiology of breast cancer in the Islamic Republic of Iran and other Asian countries such as Viet Nam (18–20)). Breast cancer was most frequently observed among women aged 36–49 years, while in western countries most of the patients were women aged 60–70 years (19,20). One reason for the lower survival rate in this study is that breast cancer was detected late compared to developed countries (19,20). A great number of round 45% of the women in our study population were newly diagnosed at Stage II of breast cancer, whereas 60–70% of the cases in developed countries were detected at lower stages (Stage I) (18,20).

The most important prognostic factor in our study was NPI. Overall survival rates at 3 and 5 years in the present study were higher than those in a similar study carried out in Viet Nam in 2013 (94% and 87% versus 83% and 74%, respectively) (18).

Tumour size was one of the strongest prognostic indicators in a previous study (9), but not in our study. A larger tumour has been associated with higher NPI and more positive lymph nodes, thus, their interaction further influences survival from breast cancer (12).

Among prognostic factors, those who were diagnosed with a poor NPI level demonstrated a poor prognosis for survival duration. This result is consistent with previous findings in a number of studies carried out in different part of the world (5–10).

The 5-year overall survival rate in our study was also higher than that obtained from research carried out in Uganda (56%), China (76.5%) and Taiwan (78.37%) (18,20,21). However, the 5-year overall survival rate was lower than that obtained in studies from the Republic of Korea (83.5%), Viet Nam (85.1%), Sweden (89%), Canada (86%) and the United States of America (88%) (18,21,22).
Roughly 2 out of every 3 breast cancers test positive for hormone receptors (23). Positivity for progesterone receptors appeared to be the most important prognostic factor for survival probability in women with breast cancer in our study. Recent studies have also shown that testing for progesterone receptors helps the doctor to choose the best treatment and reduce mortality from breast cancer. Caldarola et al. reported no significant relationship between estrogen or progesterone receptors and improved survival (24), but a later study showed patients with negative hormone receptors had a higher risk of mortality (25).

The association between skin involvement status and survival rate for breast cancer patients has been inconsistent across studies conducted around the world (5,9,10). Generally, we found a negative effect for skin involvement status on survival in breast cancer patients. The patients free of skin involvement had a lower risk of death from breast cancer than those with skin involvement.

Another prognostic factor found in this study was age at diagnosis: those who were older showed a poor prognosis for survival time. This finding is consistent with the results of previous studies (5,9,10).

The final prognostic factor in the study was nodal status: a greater number of lymph nodes reduced the survival rate among breast cancer patients in the study population; this is in accordance with previous research (26).

A number of limitations should be taken into consideration in interpreting the results of this study. The high proportion of patients censored (86.5%) may imply biased estimation in the Empirical Bayesian Additive (EBA) model. In addition, breast cancer in the study population was diagnosed when patients showed up at hospitals for primary treatment; as a result, the lead-time bias could be a potential problem.

Despite these points, health policy-makers could use the findings of this study for planning on reduction of breast cancer mortality. Among the risk factors, NPI showed a strong relationship with mortality in breast cancer. Developing early detection strategies for breast cancer in order to shift this important index to a good prognostic level is necessary in improving the life expectancy of women with breast cancer in the Islamic Republic of Iran; diagnosis has become relatively good, and the current 10-year survival rate is 76%, which is close to what is obtainable in most populations in western countries (19). Our results showed that prognostic factors of survival such as NPI, lymph node status, age, skin involvement and progesterone receptor
remain the most important determinants of 10-year survival for breast cancer patients.

**Acknowledgements**

The authors would like to thank Shiraz University of Medical Sciences and also the Center for Development of Clinical Research of Nemazi Hospital and Dr Nasrin Shokrpour for editorial assistance.

**Funding:** This study received funding from Shiraz University of Medical Sciences.

**Competing interests:** None declared.

**References**