incidence is significantly higher among males relative to females. Application of predictive model in health care is becoming an area of interest to clinicians and decision makers. As a third objective we shall use the predicted information regarding the HCC incidence to predict the future burden of the disease in terms of liver transplant costing.

The paper is structured as follows. In Section 2 we investigate the widely known risk factors for HCC and report on their distributions in Saudi. In Section 3 we construct statistical models for HCC count prediction and the relationship with the future needs for the liver transplants. A general discussion will be given in Section 4.

5LVN)DFWRUV RI +&& LQ 6DXGL $UDELD

+HSDWLWLV %DQG &LQIHFWLRQ

The most common risk factor for liver cancer is chronic infection with the hepatitis B virus (HBV) and hepatitis C virus (HCV). Individuals chronically infected with HBV and HCV are more likely to

components of what is known as δMetabolic Syndromeε, and in particular obesity, diabetes. Recent studies have highlighted that NAFLD as a chronic condition that associated with liver injury including non-alcoholic steato-hepatitis (NASH) [35]. NASH is a severe form of NAFLD and shown to be associated with obesity, insulin resistance, and mitochondrial dysfunction [36]. The prevalence of NAFLD is increasing worldwide. It is estimated that the prevalence of NAFLD in the adult US population is 30-40%, while studies from other parts of the world report a prevalence ranging from 6% to 35%. In Saudi Arabia, the prevalence of NAFLD as evaluated by computed tomography is about 10% [37]. Moreover, it has been estimated that 15-20% of patients with NAFLD have NASH. We estimate the percentage of NASH to be 5% of the NAFLD based on figures provided in reference [38]. Since obesity is one of the components of the metabolic syndrome, several studies confirmed its association with

This finding is in agreement with the fact that the risk factors NAFLD and NASH are prevalent among the middle age people. We shall restrict the following analyses to the age groups>45 years.

0RGHO ILWWLQJ

The fitted models are:

$\log \check{u}_i = L_i = \log(\text{incidence}) = \mathcal{N}_i + \mathcal{N}_i(\log(age))$ $i = 1$ for male, $i = 2$ for female

.

The fitted models are:

$L_1 = \varkappa\, 19.98 + 5.6\ (\log(age))\ R^2 = 0.915$ (1)

$L_2 = \varkappa\, 1472 + 4.1\ (\log(age))\ R^2 = 0.827$ (2)

The general relationship is thus:

$\check{u}_i = incidence\ rate = \mathcal{N}_i' a^{\mathcal{N}_i}$ (3)According, the postulated relationship between incidence and age is as given in (3) with $\mathcal{N}_i' \equiv \exp(\mathcal{N}_i)$, where $a \equiv$ age. This relationship is quite important since one can predict the burden of the disease for any particular age group based on the models (1) for males and (2) for females. The question is how good is this proposed model?

0RGHOV JRRGQHVV RI ILW

! ! wi ween inƆ bu- m' anƆ gm M(21 bu- ъ m ! tl' i ḫęlɪγlM an⌞

ꞁ

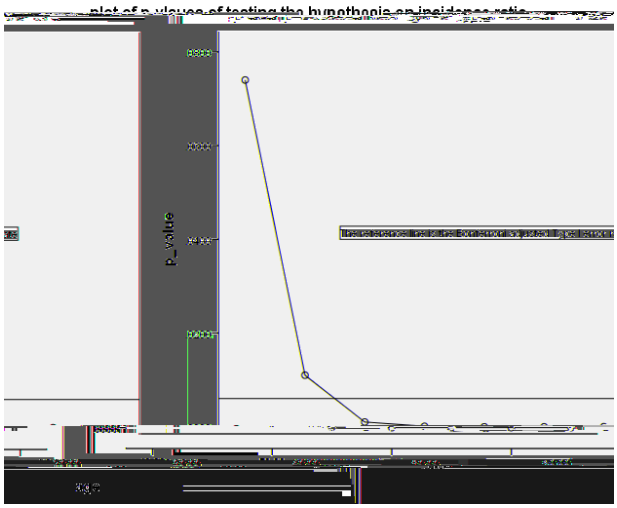

Figure 6: P-values of the statistic T for testing the departure of the incidence ratio from unity. The horizontal lines represents the p-value adjusted for multiplicity using Bonferroni's correction (p-value=0.005). The plot shows that beyond age 45 the incidence of HCC in males is significantly higher than that of females for all ages.

$E(y_{it}) = \text{var}$

3UHGLFWLQJ WKH +&& ,QFLGHQFH &RXQWV 8VLQJ 3UREDELOLW\ 0RGHOV

In this section we shall construct regression models for count (number of cases) to predict the future burden of the disease, covering a span of 10 years. Figure 7 shows the increasing trend in the number of liver cancer cases over the years for both sexes. Again, we shall restrict our analysis to ages above 35 years. For predictions, we shall employ two commonly used regression models for count data. The first is the Poisson distribution which is characterized by the equality of its mean and variance. The other model is the negative binomial distribution and is often used to model count data with over dispersion; that is when the variance is larger than the mean. In Figure 7 we show the number of liver cancer cases for males and females over the period 1997-2010.

The index of dispersion of HCC count (variance/mean) is 137 for males and 48 for females. We now describe how the models will be fitted to the data.

Since we have a time series data, we use the fixed-effects Poisson regression model for longitudinal data as described in details by Cameron et al. [44]. Let the dependent variable y_{it} denote the TB count at time t for age i (i = 1, …,n) and (t = 1,…, T). It is assumed that y_{it} have a Poisson distribution with parameter μ_{it} which, in turn, depends on a vector of exogenous variables according to the log-linear link function [51]:

$$\bar{U}_{it} = \delta_i + N\bar{y}x_{it}$$

One way to estimate the parameter of this model is to do conventional Poisson regression by maximum likelihood. For the fixed-effects Poisson regression model we have the restriction that the mean of each count must equal its variance:

model is misspecified. The closer the scaled deviance to 1, the better the fit is. Clearly both models do not provide satisfactory fit to the data. As a matter of fact, the scale parameter θ of the NB was not properly estimated. However, as suggested in ref. [52] we can tackle the over dispersion situation by multiplying the standard errors of the estimated regression parameters obtained from the fitted the Poisson distribution to the data, by the square-root of the scaled deviance and adjust the p-values upwardly. Note that in Table 2 we inflated the standard errors after multiplication by square-root of the scaled deviance=3.84, and there was no change in the corresponding p-values. The results in Table 3 justify the use of the Poisson model instead of the negative binomial even in the presence of over dispersion.

(1)	(2)	(3)	(4)
Year	**Year Code**	**HCC (x)**	**Liver Transplants (y)**

We now use the Poisson model as a predicting tool of the future

$$\hat{TLT}_n = \exp(\varkappa 0.380 + 0.00005596*NASH). \quad (10)$$

The deviance a measure of goodness of fit of the predictive Poisson model given in (10) is 1.032.

cp @ K 0 G Ð WD6 7 V G U L V N),€ 0

The predicted TLT values given in column 7 are obtained upon substitution of column 4 in the right hand side of equation (10), giving the predicted TLT values in column 7.

Now we have three sets of TLT predictions in columns 6, 7 and column 8. We examine the extent of agreement among the three sets of predictions using the intra-class correlation coefficient ρ as calculated from the one-way ANOVA [54]. We found that ρ=.997. This indicates that the TLT predicted counts whether obtained from equation (9) or equation (10) are highly correlated with each other and with the projected values given in column 6. This high agreement among the TLT predicted numbers, ensuring the robustness of the prediction models given by (9) and (10).

'LVFXVVLRQ

The etiology of HCC is extremely complex, with many confounding factors affecting the disease outcome. HBC and HCV infections increase the risk and development of cirrhosis and subsequently HCC. The development of cirrhosis and progression to HCC is complex and may involve a combination of causally interrelated factors [55]. Attempting to establish causality between HCC and the interactive effect of a host of risk factors (diabetes, obesity, HBV, HCV, age, and gender) from the Saudi cancer registry data is an impossible task. To establish causality or at least a strong association an integrated system of surveillance should be in place.

Prevention of HCC and Surveillance recommendations in adults with chronic liver disease has been established. Surveillance consists of evaluation with liver ultrasound examination, and monitoring liver enzymes every 6 months in high risk patients. Several studies have examined the effects of surveillance for hepatocellular carcinoma (HCC) in patients with liver disease [56-60]. Whereas the role of surveillance is uncertain in patients with alcoholic liver cirrhosis, studies have shown the cost-effectiveness of surveillance in patients with high risk factors. As a general rule, surveillance appears to be cost-effective in patients with cirrhosis who have an expected annual incidence of HCC exceeding 1.5 percent per year. The incidence of HCC in various forms of cirrhotic liver disease ranges from 3 to 8 percent, making surveillance cost-effective for many patients with cirrhosis. For patients with hepatitis B, surveillance appears to be cost-effective, once the incidence of HCC exceeds 0.2 percent per year [61-63].

In our study no attempt was made to link HCC with its risk factors, since we have no record or registry of the incidence of HBV and HCV infection, nor the other risk factors. In Saudi Arabia, there is no surveillance program for patients with chronic liver disease, and the expatriates represent around 30% of the population. In an attempt to reduce the prevalence of infection and its potential sequelae including HCC, the health authorities established strict regulations for HBV and HCV screening for all expatriates applying for working visas. The regulations prohibit issuing work license or visas for infected persons or those with carrier status. The authorities also developed a national immunization programme for HBV. These undertaken measures were an important component of successful control of HBV and HCV infection. Also, there has been increased awareness of the HBV and HCV infection in the country, and increasingly effective prevention and treatment are becoming available to treat the infection and reduce the complications. The authorities have also launched awareness campaigns for healthy life style to reduce obesity and diabetes among the Saudi population.

In Saudi Arabia, there is no surveillance program for patients with chronic liver disease. In an attempt to reduce the prevalence of infection and its potential consequences authorities, developed a national immunization program for HBV. These undertaken measures were an important component of successful control of HBV and HCV infection. Also, there has been increased awareness of the HBV and HCV infection in the country, and increasingly effective prevention and treatment are becoming available to treat the infection and reduce the complications.

Recently, it has become evident that there is a significant association between obesity, and NAFLD, leading to NASH and possible liver transplant. Experts [64,65] argue that the sooner the body weight is reduced the healthier the liver will become. In summary, the growing body of research on NAFLD and NASH seems to suggest that obesity, diabetes, or may be all the components of metabolic syndrome may explain the increasing prevalence of NAFLD. Recent studies [63] showed that lifestyle interventions have substantial health benefits. The trend in BMI is on the rise in Saudi Arabia [39]. The interesting summary in [55] and the references therein it is recognized that the etiology of HCC is extremely complex, with multiple many confounding factors affecting disease course and patient prognosis. For example, most patients with HCC have cirrhosis, which develops following long periods of chronic liver disease. Moreover both HBV and HCV infection increase the likelihood of developing liver cancer, with an incidence of 54.4% and 31.1% of liver cancer cases globally, respectively [55]. Patients with NAFLD/NASH, also appears to further drive the hepato-carcinogenic process [55]. It is not clear how exactly role of each of these pathways and this is due in part to the lack of available data. Further research efforts are therefore needed to fully elucidate the diverse mechanisms involved in the pathogenesis of HCC.

Therefore fiscal prudence by policymakers, makes intervention to reduce obesity and hence the risk of developing NAFLD, a desirable strategy.

The fundamental aim of this paper is to establish a reliable statistical model through which projections for future HCC counts and the needed number of liver transplants are provided. We have used the well-known Poisson regression model, and estimated its predictive accuracy to achieve these objectives. Once this is done, an estimate of the total cost of transplants would be readily obtained.

The decision-making process in many aspects of the healthcare delivery system is often complicated and detailed, requiring many considerations before arriving at a course of action in patient care. Various modeling tools have been devised to assist with the decision-making process. Some are aimed at predicting a clinical outcome, whereas others focus on the cost of care.

For example several studies used and applied predictive modeling in health care [66]. They used logistic regression model, which is a member of the generalized linear models family to prospectively predict the total annual health costs (medical plus pharmacy) of health plan participants using Pharmacy Health Dimensions (PHD).

There is no doubt that our decisions would be infallible if we had reliable data that provided an absolute clear outcome for decision-

making [67,68]. The important question that needs to answered; is whether predictive modeling is useful in the short term? Could the predictive model lead us to making decisions that benefit us in the short-term? Is our future success found somewhere in the success of yesterday? Before we answer these questions we should agree that, if we have very little data, then we should opt for a simple prediction model like the generalized linear regression (Poisson regression) model that we used in this paper. It was important to evaluate a possible competitor namely the negative binomial regression model. This letter was not as good as the former.

Many factors must be considered in the development of predictive models to ensure their validity. The model should be user-friendly, robust and flexible enough to handle the dynamic nature of the factors affecting the prediction of the outcome of interest.

With costs today heavily influencing benefit decisions of employers and other payers, the ability to predict outcomes becomes all the more important.

Predictive modeling has become a popular way to assess and manage costs associated with healthcare. The core δcompetencyε of predictive modeling is its proactive nature; that is, not taking a reactive approach to disease management and cost containment but instead identifying trends and forecasting events that may have substantial implications for healthcare.

When using prediction modeling to Saudi population data, key principles should be considered. Firstly the dynamic nature of this emerging economy, secondly need for large disease data base and an integrated surveillance system, and finally a supportive decision making system. The decision-making process in many aspects of the healthcare delivery system is often complicated and detailed, requiring many considerations before arriving at a course of action in patient care. Various modeling tools have been devised to assist with the decision-making process. Some are aimed at predicting a clinical outcome, whereas others focus on the cost of care. As this research area progresses, it is important to not just show the academic value of these models for predictive accuracy but also to provide evidence of benefit to the many healthcare stakeholders.

In the USA the cost of a transplant, including preliminary testing, the surgery itself and post-operative recovery costs vary across the country and depend on the hospital and organ type. These costs start to add up, even prior to the transplant. In fact, patients commonly rely on several sources to help pay for their medical and non-medical costs of pre- and post-transplantation. It was reported in 2010 that the average cost of liver transplant in the USA can be as high as $577,100. The cost covers $25,400 for 30 days pre-transplant, $71,000 procurement, $316,000 hospital transplant admission, $46,600 physician during transplant, $93,900 180 days post-transplant

56. Saab S, Ly D, Nieto J, Kanwal F, Lu D, et al. (2003) Hepatocellular carcinoma screening in patients waiting for liver transplantation: a decision analytic model. Liver Transpl 9: 672.
57. Sherman M, Peltekian KM, Lee C (1995) Screening for hepatocellular