Efficacy and Effectiveness of Liver Screening Program to Detect Fatty Liver in the Periodic Health Check-Ups

Kyoko Nomura¹, Eiji Yano¹, Toshiaki Shinozaki¹ and Kazumi Tagawa²

¹Department of Hygiene and Public Health, Teikyo University School of Medicine and ²Department of Internal Medicine, Mitsui Memorial Hospital, Japan

Abstract: Efficacy and Effectiveness of Liver Screening Program to Detect Fatty Liver in the Periodic Health Check-Ups: Kyoko Nomura, et al.—To determine whether the current liver screening program for fatty liver has sufficient scientific evidence to justify its continued implementation. The liver screening program to detect fatty liver was performed on 411 Japanese workers utilizing serum aspartate aminotransferase (ALT), alanine aminotransferase (AST), and γ-glutamyl transpeptidase (γ-GTP). Based on the preceding studies, subjects with viral and alcohol hepatitis were excluded from the evaluation. The diagnosis of fatty liver was based on ultrasound findings. The program was evaluated by efficacy and effectiveness; efficacy was measured according to the receiver operating characteristic (ROC) curves in comparison with the Body Mass Index (BMI). Effectiveness, based on the efficacy determinations, was assessed by means of the positive predictive value (PPV) test performance, the disease characteristics, and the program price. The diagnostic performances of ALT and BMI were nearly acceptable but far from excellent. The areas under the curves of the two indices were 0.69 and 0.63, respectively and these were statistically equivalent. The PPV ranged from 15 to 28% where the prevalence of fatty liver was 12.3%. The price of the program was estimated at US 4 dollars per person based on the medical reimbursement fee rate. The efficacy of the liver screening program was found to be insufficient and BMI monitoring may provide a more suitable and inexpensive alternative. Furthermore, the effectiveness of the program is open to question, considering the generally benign prognosis of the disease in the absence of any accompanying morbid conditions and the high price of the program. (J Occup Health 2004; 46: 423–428)

Key words: Effectiveness, Efficacy, Fatty liver, Periodic health check-ups, ROC

In 1989 the mandatory program involving the annual health check-ups of the main target population of workers aged 40 and over was amended in Japan under the Occupational Safety and Health Law. This includes the various ordinances, such as medical examination at the time of employment (the Article 43), periodical medical examination (the Article 44) and medical examination for those engaged in designated work (the Article 45) and emphasizes the responsibility of an employer to provide such medical check-ups by the physician for regular employees. Among these ordinances, we focused on workers who undergo the periodic medical check-up at least annually, namely subjects under Article 44. Since there are 60 million workers in Japan and those aged between 40 and 65 yr account for 34% of the Japanese population¹, approximately 20 million people are mandated for check-ups with liver function test every year, which is an expensive preventive health practice. The check-ups include various screening programs, but, according to an assessment of clinical preventive interventions in the United States², Canada³, and the UK⁴, the evidence for the inclusion of some specific diagnostic tests within the programs is equivocal. These authorities advised that a screening test should not be performed where the target population or purpose is unclear. As the cost of the periodic health check-ups is paid for by the companies, they impose a very large financial burden on industry. Mindful of the scarcity of health care resources, it is imperative that every screening program enforced by law should be founded on valid scientific evidence.

The liver screening program in the check-ups was evaluated by measuring three serum hepatic enzymes:
(aspartate aminotransferase: AST, alanine aminotransferase: ALT, and gammaglutamyl transpeptidase: γ-GTP). The Labor Standard Office Report for the check-ups in 2003 showed that liver dysfunction yields the second highest abnormality rate (15.5%) after hyperlipidemia (28.4%). The WHO has listed the basic principles and four guideline conditions for the introduction of effective screening programmes. Namely: 1) the disease poses an important public health problem, 2) the natural history of the disease is well understood, 3) the disease should present with a latent or pre-symptomatic phase, 4) the disease process should be reversible or more amenable to therapeutic or ameliorative intervention during its latent or pre-symptomatic phase.

With reference to the WHO guidelines, and excluding acute and rare diseases, and also non-alcoholic steatohepatitis (NASH), three chronic liver diseases are encountered as target diseases in the current liver screening program. Namely, infection with hepatitis B and C virus (HBV and HCV respectively), alcoholic hepatitis and fatty liver. In previous studies, we have investigated HBV and HCV, and alcoholic hepatitis, and have made the following two suggestions: (1) The proposed introduction of anti-HCV antibody screening once in a lifetime is more cost-effective than the annual compulsory liver function tests for serum ALT, AST and γ-GTP. (2) The efficacy of the current liver function tests to detect alcoholic hepatitis exhibits very low sensitivity, and therefore questionnaires to monitor alcohol intake may be more suitable.

In the present study, fatty liver has been further investigated since it has the highest prevalence among the target diseases in the liver screening program, and hence the purpose of the screening could be more precisely formulated.

Thus, the detection program for fatty liver was evaluated at two levels: efficacy of the liver function tests used in the program to detect fatty liver, and also effectiveness in the actual occupational setting where the screening program is performed. The first level was measured with the receiver operating characteristic (ROC) curves in comparison with an alternative simple and inexpensive screening index, namely Body Mass Index (BMI). Secondly, based on the efficacy, effectiveness was further assessed by the positive predictive value (PPV) of the tests, the disease characteristics, and the price of the program.

Methods

Study subjects

The investigation was carried out on 411 Japanese workers aged between 18 and 61 during the course of periodic health check-ups in a metal-working company in 1999. For the purpose of the study, those who met the following criteria were excluded from this study. (1) hepatitis virus carriers diagnosed by HBs antigen positive (n=12); (2) HCV antibody positive (n=7); (3) excess alcohol drinkers with habitual daily alcohol intake comparable to or more than 540 ml of Japanese sake (n=48); (4) female workers (n=21). In defining the target disease, our previous studies showed the following two points: (1) an anti-HCV antibody measure would be the most cost-effective strategy compared to either no examination strategy or the currently implemented serum test. Moreover, because the disease has a poor prognosis, the screening program for viral hepatitis should not rely on a serum liver function test which exhibits a high false negative rate. (2) with regard to alcoholic hepatitis, a questionnaire directly asking subjects for their drinking history yields higher efficacy than that of a serum liver function test. Accordingly, the target disease in the present study was defined as fatty liver due to over-nutrition. Analyses restricted to fatty liver due to neither alcohol nor viral infection would make an unbiased assessment possible. Therefore the excess alcohol users and hepatitis virus carriers were excluded. Since three subjects were both HBs antigen and HCV antibody positive, a total of 326 workers qualified as participatory subjects for analysis in this study. The median age in an interquartile range was 47 (35, 52) years.

The following measurements were undertaken in each individual: three hepatic enzymes namely ALT, AST, and γ-GTP, and also BMI. These in plasma were determined with an automatic biochemical analyzer (Olympus AU-5400, Japan). The biochemical assay was performed by a commercial laboratory. Quality control was assessed by measuring the reference materials and depiction on a quality control chart. BMI is a measure of weight in relation to height: BMI = weight (kg)/height² (m). Following the recommendation by WHO and the Japan Society for Obesity, obesity was defined as a BMI ratio greater than 25 kg/m². Among the subjects, 63% had a BMI less than 25, and 10% had hypertension defined as systolic blood pressure greater than 160 mmHg, and another 10% had hyperglycemia defined as blood sugar greater than 120 mg/dl. Subjects having hyperlipidemia defined as total cholesterol greater than 234 mg/dl were 25%. For smoking and drinking habits, current (past) users were 60% (21%) and 80% (19%), respectively.

The diagnosis of fatty liver was based on ultrasound examination. The ultrasound examiners were not informed of the purpose of the present study, nor knew the results of the serum liver function test. The fatty liver can be characterized by the high echo intensity of the liver parenchyma of the ultrasound image. An equivocal finding was considered to indicate not having fatty liver. The ultrasound is not a primary screening examination and is generally regarded as a diagnostic test performed for those whose liver function screening tests are positive. But, in this particular factory, it was
performed on all the check-up participants.

**Evaluation of liver screening program in fatty liver**

**Efficacy**

Based on the discriminatory thresholds currently used in a check-up of AST (<39 IU), ALT (<36 IU), γ-GTP (<61 IU) and BMI (<25 kg/m²), sensitivity, and specificity were calculated. The 95% confidence intervals (95%CI) were also calculated by using the exact method. The Youden Index\(^\text{14}\) was used to give the best cut-off points for each screening index, and based on the indices, the test performances were again calculated with 95% CI. The area under the ROC curves (AUC) was calculated by the Delong method\(^\text{15}\) in each liver enzyme test and BMI. According to Hosmer and Lemeshow, AUC between 0.7 and 0.8 is considered as “acceptable” and between 0.8 and 0.9 as “excellent” discrimination\(^\text{16}\).

**Effectiveness**

Referring to the previously reported requirements concerning the evaluation of screening programmes\(^\text{7, 17}\), the effectiveness of a liver screening program to detect fatty liver in this study was evaluated for the following two aspects and the related aspect was discussed in the Discussion section.:

1. When the target disease of the screening program is relatively common among the target population, a larger number of persons with true fatty liver will be detected among the test positive. To demonstrate this, the PPV test performance was exploited.

2. The screening test should be inexpensive, non-invasive, acceptable to the test population, and exhibit high efficacy. The test is performed by blood drawing & collection. Information with regard to the price of the tests was also gathered.

**Statistical analysis**

Analyses were conducted with the Intercooled Stata, version 7.0 for Windows and the SAS Version 8.12 for Windows. A \(p\) value of less than 0.05 was considered significant.

**Results**

Among 326 examinees, subjects with an abnormal range of the three enzyme indices ranged between 12 and 20%, and those with one or more abnormal indices were 30%. One hundred and twenty subjects (37%) were obese with BMI =>25 kg/m² and 40 subjects (12.3%) had fatty liver.

Table 1 shows screening indices to detect fatty liver in biochemical markers and BMI. The sensitivities of AST and γ-GTP were very low under the cut-off points currently used in the health check-up (18% and 20%, respectively) and remained

### Table 1. Screening indices to detect fatty liver in biochemical markers and BMI (n=326)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>AST (IU) 95%CI</th>
<th>ALT (IU) 95%CI</th>
<th>γ-GTP (IU) 95%CI</th>
<th>The three enzyme combination 95%CI</th>
<th>BMI (kg/m²) 95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut-off points in current check-up</td>
<td>39 (12%)</td>
<td>36 (20%)</td>
<td>61 (16%)</td>
<td>98 (30%)</td>
<td>120 (37%)</td>
</tr>
<tr>
<td>Abnormal range N (%)</td>
<td>38 (12%)</td>
<td>64 (20%)</td>
<td>51 (16%)</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>17.5</td>
<td>7.3–32.8</td>
<td>45.8–77.3</td>
<td>20</td>
<td>9.1–35.7</td>
</tr>
<tr>
<td>Specificity</td>
<td>91</td>
<td>87.0–94.0</td>
<td>78</td>
<td>72.7–82.6</td>
<td>85</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>88.7</td>
<td>84.6–92.1</td>
<td>93.7</td>
<td>89.8–96.4</td>
<td>88.4</td>
</tr>
<tr>
<td>Best Youden index*</td>
<td>0.13</td>
<td>0.48</td>
<td>0.19</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Cut-off point of the best YI</td>
<td>21</td>
<td>30</td>
<td>41</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>87.5</td>
<td>73.2–95.8</td>
<td>82.5</td>
<td>67.2–92.7</td>
<td>45</td>
</tr>
<tr>
<td>Specificity</td>
<td>25.2</td>
<td>20.3–30.6</td>
<td>65.4</td>
<td>59.6–70.9</td>
<td>73.8</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>14.1</td>
<td>10.0–19.0</td>
<td>25</td>
<td>17.9–33.3</td>
<td>19.4</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>93.5</td>
<td>85.5–97.9</td>
<td>96.4</td>
<td>92.7–98.5</td>
<td>90.6</td>
</tr>
<tr>
<td>AUC</td>
<td>0.53</td>
<td>0.47–0.59</td>
<td>0.69</td>
<td>0.61–0.77</td>
<td>0.51</td>
</tr>
</tbody>
</table>

*Youden WJ. Index for rating diagnostic tests. Cancer,1950
CI: Confidence Interval, AST: aspartate aminotransferase; ALT: alanine aminotransferase; γ-GTP: gammaglutamyl transpeptidase; YU: Youden index
The efficacy of ALT, BMI and the three-enzyme combination were moderately good and equivalent. Similarly, ROCs for these indices approached acceptability, but were far from excellent: ALT (AUC:0.69; 95% CI, 0.61–0.77) and BMI (AUC:0.63; 95% CI, 0.56–0.70).

The comparison between each of the three liver enzymes and BMI, ROC curves are shown in Fig. 1. The AUC significantly favors BMI over both AST and γ-GTP (p=0.041 in Fig. 1-a and p=0.023 in Fig. 1-b, respectively) but ALT and BMI were equivalent (p=0.244 in Fig. 1-c).

Among efficacy indicators, PPV remained low, ranging from 16 to 28%. Using even the best cut-off points and Youden index, PPV did not increase because of the low prevalence of fatty liver in the present study population (Table 1).

The assay price of the liver screening program was incorporated into the total price of the periodic health check-ups but, a commercial laboratory source revealed that the assay price was based on the medical reimbursement rate. Accordingly, the rates were: GOT and GPT cost 220 Japanese Yen (JPY) for each and γ-GTP costs 130 JPY, which jointly account for approximately US 4.75 dollars per person (US1$=120JPY) for the liver screening program.

**Discussion**

The results of the present study show that the efficacy of the current liver screening program was not sufficient to detect fatty liver even using the best Youden index. The absence of any significant difference in AUC between ALT and BMI suggests the possibility of using BMI as an inexpensive alternative index. The PPV test performance was limited because of the low prevalence of fatty liver in the target population. That is, the numbers of true fatty liver detected among the test positive would be limited and therefore the effectiveness of the test is questionable. The natural history of fatty liver is generally benign in the absence of accompanying morbid conditions and preexisting fibrosis and steatohepatitis (NASH). Furthermore, in the current check-ups, various screening programs targeting on these accompanying conditions are already implemented. In addition, the overall cost of the screening program, involving approximately 20 million eligible workers for the annual check-ups, required a large amount of money.

A number of limitations to the interpretation of the results of the present study are acknowledged: first, in the generalizability of the results, since the majority of our study population had a normal BMI of less than 25 (kg/m²). Obesity is a multi-factorial condition and the combination of various cultural, economic and social parameters are known to play a complex causal role.
In this study, the sample population was all blue-collar workers in a metal-working company. Therefore reservations might apply when extrapolating the results of the current study to other populations such as young white-collar workers with a higher prevalence of fatty liver.

Second, for a definitive diagnosis of fatty liver, liver biopsy exhibits a higher sensitivity than that of ultrasound, but, because the former is invasive and expensive, ultrasound is more commonly used in Japan. Indirect visual diagnosis of fatty liver can introduce misclassification but its magnitude might not be so large. The prevalence of fatty liver in this study was about equal to that of a much larger study which targeted Japanese workers in general.

Based on the result of the present study, an appraisal of the effectiveness of the current liver screening program to detect fatty liver according to the two conditions proposed in the method section, is presented.

(1) The study shows that under the circumstances in which a target disease is not common in the population, PPV test performance is limited. Fatty liver accompanying such diseases as hypertension, hyperlipidemia and diabetes are all related to obesity. In order to increase the effectiveness of the current program in fatty liver, restriction of persons thought to be at “high risk” may be a possible and preferable alternative approach in screening for this disease.

(2) Although blood sampling is a simple medical procedure and acceptable to the test population in Japan, the present study indicates that the high financial cost of implementation means that cost-effectiveness and cost-benefit issues are of major concern in relation to health economic and prudent political decision making.

Since, in the absence of clinical complications (e.g. pre-existing fibrosis and non-alcoholic steatohepatitis, NASH), fatty liver is a benign disease, the detection of fatty liver per se is of questionable medical significance. In this study, NASH was not assessed because the prevalence of NASH is low and not fully understood pathophysiologically. Furthermore, research suggests that the entire histologic spectrum of NASH can be seen in individuals with normal ALT values, and a low normal ALT value does not necessarily indicate freedom from underlying steatohepatitis with advanced fibrosis. If this disease were to be included in the program, subjects would participate repeatedly in the same screening program but without achieving a diagnosis.

In conclusion, the efficacy of the current liver function tests to detect fatty liver was found to be far from excellent, and furthermore, of dubious effectiveness for the population receiving periodical health check-ups. As mentioned earlier, based on the WHO monograph, after excluding acute and rare diseases, the target diseases for the current liver screening program are infection with hepatitis B and C virus (HBV and HCV respectively), alcoholic hepatitis and fatty liver. Our previous two studies revealed that alcohol history taking had higher efficacy than that of serum hepatic examination, and an anti-HCV antibody measure was the most cost-effective strategy compared to either no examination strategy or the currently implemented serum test. Moreover, in general more than 80% of those with HB antigen were healthy carriers with no abnormality in liver enzymes, so that those with hepatitis B would not gain any benefit from a serum hepatic test. Hence, in the absence of substantive scientific evidence and demonstrable potential benefit, support for the continued implementation of the costly liver screening program is thereby considerably weakened.

**References**

18) Interpretation of Medical Reimbursement Rate [Ikatensuhyou No Kaishaku]: Social Insurance Research Institute, [Shakai Hoken Kenkyujo], 2003 (in Japanese).