Evaluation of Current Guthrie TSH Cut-off Point in Iran Congenital Hypothyroidism Screening Program: A Cost-Effectiveness Analysis

Ahmad Reza Shamshiri MD\(^1\), Shahin Yarahmadi MD\(^2\), Mohammad Hossein Forouzanfar MD PhD\(^3\), Ali Akbar Haghdoost MD PhD\(^4\), Gholamreza Hamzehloo MLTD\(^5\), Kourosh Holakouie Naieni PhD\(^1,6\)

Abstract

**Background:** The threshold of thyroid-stimulating hormone (TSH) in current screening for congenital hypothyroidism (CH) from the heel prick test is 5 mU/l. This study uses cost-effective analysis to evaluate increasing the threshold to minimize false-positive results and recall rates.

**Methods:** Cost of screening, diagnosis and treatment, education, and care of mentally retarded patients were gathered from the Ministry of Health State Welfare Organization and Department of Education in Tehran. Sensitivity analysis and calculation of confidence interval for incremental costs and effects (gained disability adjusted life years – DALYs) and incremental cost-effectiveness ratios (ICER) were performed by Monte Carlo simulation with Ersatz software.

**Results:** ICER for screening programs with different TSH cut-off points versus no screening was similar, and approximately -4.5 ± 0.2 thousand US dollars per gained DALY. In the proposed cohort (10,000 neonates), gained DALYs were 316 ± 50 for a cut off point of 5 mU/l, 251 ± 40 for 10 mU/l, 146 ± 23 for 15 mU/l, and 113 ± 18 for a cut-off point of 20 mU/l. Sensitivity analysis showed that the model remained the same when the input parameters were changed.

**Conclusion:** This study demonstrates that the current threshold of TSH in the national CH screening program in terms of cost-effectiveness is the most appropriate threshold. However, more studies are needed to examine new strategies and methods to reduce recall rates and related consequences such as repeated thyroid testing in neonates.

Keywords: Congenital hypothyroidism, cost-benefit analysis, neonatal screening, quality-adjusted life years, thyrotropin


Introduction

Congenital hypothyroidism (CH) is the most common preventable etiology of mental retardation, which is hardly diagnosed due to late presentation of signs and symptoms, unless screening is performed. For this reason, many countries that have resources and possibilities for CH screening include this plan in their neonatal screening package.

In Iran, CH screening with umbilical cord sample was established before the salt iodization program in 1987, and then stopped due to high recall rate.\(^1\) The pilot of the new screening program with the heel prick testing was launched in 3 provinces in 1997 and integrated into the National Health System in 2005.

New studies\(^2\) have reported that the incidence of CH disease is higher (1.1 – 2.4 in 1,000 births) in Iran when compared to the frequency statistics of disease worldwide (1 in 3,000 births).\(^5\) It means that we are confronted with 2,700-3,000 new patients per 1,300,000 births annually.\(^6\)

Among the different methods of neonate screening, Iran has selected the protocol of CH screening with heel prick test and thyroid-stimulating hormone (TSH). According to the general rule, there is no defined cut-off point for such a screening program. Therefore, each country should start screening with a conservative cut-off point and collect enough data to make the final decision for an optimum threshold with the intent to reduce recall and false positive rates. In contrast to diagnostic tests, we have accepted some false positives, but attempted to achieve a false-negative rate close to zero in our screening tests.\(^7\) For this reason, in addition to some unpublished studies on CH patients as well as their screening TSH levels, the Guthrie cut-off point has been adjusted to the minimum possible value of 5 mU/l in Iran.

It is clear that with the changing pattern of this disease and improved laboratory techniques, this cut-off point should be updated periodically. The initial cut-off point of 15 mU/l was chosen for TSH in Canada, but since 2001 the threshold for borderline cases in the second test has been decreased to 5 mU/l. This decrease has resulted in more cases being detected.\(^7\) Changing the program in the US also resulted in finding duplicate cases from 2001 – 2004 in contrast to 1991-1994.\(^4\) This increase mainly accounted for mild and delayed cases that needed to be treated.

In the health section, the lack of resources and the manner of re-
source allocation to different interventions is always a major challenge. Analysis of cost-effectiveness is a common approach to determining the value of interventions, and refers to a method for assessing their costs and health benefits. In this analysis, monetary and health outcomes are measured separately and the relative value of an intervention is measured as the additional cost it takes to achieve an incremental health benefit (such as dollars to prevent a case of mental retardation). Meanwhile, the cost-effectiveness of the CH screening program has been accepted worldwide and is known as a cost-saving intervention.9 While there are studies available on the cost analysis of the program in Iran10–11 and other countries,12–15 residual ambiguities in our selected TSH screening threshold remain. There is a question whether we can raise the cut-off point and consequently decrease the recall rate and laboratory referrals while still maintaining program efficiency. Thus in this study we attempt to examine the current CH screening program with the Guthrie TSH cut-off point of 5 mU/l against the upper thresholds through cost-effectiveness analysis (CEA).

### Materials and Methods

In this study, we defined four screening programs as interventions, each with different Guthrie TSH cut-off points. In the following sections, we show how cost and health benefit estimations for each screening program have been performed and how CEA has been performed against a no-screening situation in a hypothetical cohort of newborns.

#### National CH screening program

The protocol of the current program for CH screening in Iran is the heel prick test of TSH in neonates that are 3 – 5 days old. Due to a high frequency of CH in Iran, and to minimize false negatives, the TSH cut-off point is set at 5 mU/L. Neonates with a Guthrie TSH test equal to or more than 5 mU/L are recalled for more confirmatory tests (with venous blood samples) and then a final decision is made by each medical university focal point (pediatric endocrinologist).16

#### Tehran CH screening program data

In 2009, there were 34,007 CH screenings performed by the Central Health Laboratory of Tehran University of Medical Sciences, which covered 2,700,000 people. Of these, 2,310 (6.79%) neonates who had heel prick tests of TSH ≥ 5 mU/l were recalled for further thyroid function tests in 3 other laboratories that used the chemiluminescence immunoassay technique. The results from testing 1,197 neonates who were referred to the aforementioned laboratories revealed that 39 neonates had CH disease using the laboratory criteria of TSH ≥ 10 mU/L and total T4 ≤ 6.4 ng/dl. Unfortunately, clinical data of neonates screened for CH were not accessible for final clinical diagnosis. However, consistent with previous studies, the incidence of CH in Tehran was estimated to be 2.2 per 1,000 births.

#### Costs

In this study, the costs were assessed from the caregiver’s point of view. The cost of screening, the confirmatory test, general and specialized physician visits, drugs, periodic laboratory tests, education, and care of mentally retarded patients were derived from the Yarahmadi et al. study.10 The main sources for cost estimation were the Ministry of Health and State Welfare Organization in 2008, when the exchange rate was about 1 U.S. Dollar to 10,000 rials. We assumed that 80% of patients had health insurance, but the cost of screening services was not covered by health insurers.

Regular education in public schools is free of charge. The average cost of education in private schools of Tehran was derived from the website of Department of Education in Tehran17 by a random selection of 10 schools from different categories, divided into sex and educational levels. In Tehran, about 12% of children study in private schools.

For special education and care of mentally retarded patients, there are both public and private centers. The patient care centers are also divided into day-time and 24-hour centers. The exact pattern of using these centers by patients’ families is unknown. Yarahmadi et al.10 has defined 6 scenarios in their study for patients’ care: (a) all patients in public day-time centers, (b) all patients in private day-time centers, (c) all patients in public 24-hour centers, (d) all patients in private 24-hour centers, (e) equal ratio in public and private day-time centers, and (f) equal ratio in public and private 24-hour centers. In reality, all of these scenarios are not possible. Thus, we considered an alternative scenario for the care of mentally retarded patients: 25% in public daytime centers, 25% in private daytime centers, 25% in public 24-hour centers, and 25% in private 24-hour centers. In this study it was assumed that half of the patients attended private schools, while the remainder attended public schools.

### Table 1. Lifetime discounted cost (US dollars) and effectiveness parameters of congenital hypothyroidism (CH) screening with four different Guthrie TSH cut-off points in a theoretical cohort of 10,000 newborns.

<table>
<thead>
<tr>
<th></th>
<th>CH* screening with cut-off = 5 mU/l</th>
<th>CH screening with cut-off = 10 mU/l</th>
<th>CH screening with cut-off = 15 mU/l</th>
<th>CH screening with cut-off = 20 mU/l</th>
<th>No-screening for CH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of Guthrie test</td>
<td>24,196</td>
<td>24,196</td>
<td>24,196</td>
<td>24,196</td>
<td>0</td>
</tr>
<tr>
<td>Cost of confirmatory tests</td>
<td>4,507</td>
<td>840</td>
<td>271</td>
<td>139</td>
<td>0</td>
</tr>
<tr>
<td>Cost of diagnosis†</td>
<td>14</td>
<td>44</td>
<td>92</td>
<td>107</td>
<td>159</td>
</tr>
<tr>
<td>Number of CH patients by screening</td>
<td>20/22</td>
<td>16/22</td>
<td>9/22</td>
<td>7/22</td>
<td>0/22</td>
</tr>
<tr>
<td>Cost of laboratory tests</td>
<td>6,279</td>
<td>6,279</td>
<td>6,279</td>
<td>6,279</td>
<td>6,279</td>
</tr>
<tr>
<td>Cost of visits by physicians</td>
<td>6,888</td>
<td>6,888</td>
<td>6,888</td>
<td>6,888</td>
<td>6,888</td>
</tr>
<tr>
<td>Cost of drugs</td>
<td>1,161</td>
<td>1,161</td>
<td>1,161</td>
<td>1,161</td>
<td>1,161</td>
</tr>
<tr>
<td>Cost of education</td>
<td>17,056.491</td>
<td>17,120,516</td>
<td>17,224,557</td>
<td>17,256,569</td>
<td>18,878,927</td>
</tr>
<tr>
<td>Cost of care</td>
<td>11,833</td>
<td>252,128</td>
<td>642,608</td>
<td>762,756</td>
<td>1,286,166</td>
</tr>
<tr>
<td>Healthy life years</td>
<td>263,003</td>
<td>262,938</td>
<td>262,832</td>
<td>262,800</td>
<td>262,686</td>
</tr>
</tbody>
</table>

* CH = Congenital hypothyroidism; † Cost of diagnostic tests for CH patients when no screening implemented in the population or screening failed to discover a case.
Figure 1. Decision tree of congenital hypothyroidism. * CH: Congenital hypothyroidism, ** MR: Mental retardation

Figure 2. Cost-effectiveness acceptability curve of congenital hypothyroidism screening programs with different Guthrie TSH cut-off points.

Figure 3. Cost-effectiveness plane of congenital hypothyroidism screening programs with different Guthrie TSH cut-off points. (To show overlapping data points, drawings are manually applied around them.)
In addition to treatment costs, indirect costs related to the non-productivity of the patients, as well as additional costs that impose financial burdens on families, were significant. The burden of stress and psychological tension that imposed on neonates and their families who participated in the screening program, particularly those recalled for confirmatory tests, should be considered. However, due to the complexity in estimating costs, they were ignored in this model.

CH screening effectiveness

The effectiveness of the CH screening program was quantitatively calculated by gained disability adjusted life years (DALYs). For a cohort of neonates, the disability weight of CH was derived from the national burden of diseases (2003) that was similar to the disability weight of Dawn syndrome, and was equal to 0.593 for all age groups. There is no evidence that CH patients have a different pattern of mortality in comparison with the general population. The recent mortality rates in Iran (by gender and 5-year age groups) were obtained from the Ministry of Health (un-published data) and applied to both CH patients with or without mental retardation, and other people without CH.

As CH patients like other people might coincidentally exposed to different disabling diseases and conditions, we adjusted the disability weight of CH disease with age-specific weighted disability prevalence estimated for the Iranian population by a multiplicative model. Also, we applied the same weighted disability prevalence for unaffected and treated patients (without mental retardation) in the CEA model.

CEA model and parameter assumptions

According to statistics released in 2009, the CH screening program was supposed to be equal to 92%, but we know that the real coverage has been extended by the addition of screening tests in private laboratories. The incidence of CH disease was considered equal in males and females. The number of annual physician visits, the amount of drug administrations, and annual laboratory tests were extracted from the Yarahmadi et al. study. Treatment duration for transient cases of CH was assumed to be 3 years. Also, we assumed that all symptomatic neonates would be diagnosed as mentally retarded and all patients found in the screening program would be treated properly, and would have a normal intelligence score.

According to unpublished data from the Ministry of Health (Endocrine and Metabolic Office, Center for Disease Control), the false-negative rate was 1% and the transient CH rate was considered 0.31.

Discount rate

As with the majority of previous studies, the same discount rate of 3% has been used for costs and effects associated with a screening program.

Analysis of CEA

A decision tree that demonstrated the screening versus non-screening outcomes was drawn according to the natural history of CH disease (Figure 1). The accuracy of the model was checked and approved by replacing different input parameters with a range of values and a possibility of outcomes.

The simulation model was carried out using a cohort of 10,000 neonates (5,000 females and 5,000 males) from birth to a maximum of 82 years of age. At first, the number of permanent and transient cases of CH and the proportion of those who were discovered by applying the program coverage and false-negative rates were calculated. The number of annual deaths was estimated year by year, according to the national mortality rate. For calculating the living years, we assumed that all deaths occurred in the middle of year except deaths that occurred before 1 year, which we considered they lived for 0.1 years, and before 2 years, which we considered they lived for 1.4 years. Then, the life years of mentally retarded patients and other populations were adjusted for the disability weight of CH and age-specific weighted disability prevalence, as described above. In this manner the gained DALYs were calculated for the cohort.

Considering the disease incidence rate and screening program coverage, the total cost of screening and confirmatory laboratory
tests were calculated. The cost of physician visits, drugs, and laboratory tests were also calculated year by year for patients. Therefore, we considered the cost of education applied for the cohort during their school ages and the cost of care of mentally retarded patients until the end of their lives. The total cost for the cohort was thus calculated.

The model was accomplished with cut-off points of 5, 10, 15, and 20 mU/l. For each scenario, the incremental costs of screening versus no-screening were divided by incremental life years without disability (gained DALYs) and the ICER was calculated. ICER shows how much costs have been paid for each gained DALY.

To calculate a 95% confidence interval (CI) for costs, effects, and ICER, Ersatz software (http://www.epigear.com; Monte Carlo add-in program for MS Excel) with 2,000 iterations was used. Cost-effectiveness plane and acceptability curves were drawn by putting outputs of the software into Excel.

Sensitivity analysis

In Ersatz software, we can define specific distributions for input parameters to take uncertainty around them. We considered a Poisson distribution for diagnosed patients using a CH screening program and gamma distribution for all costs of education and care of mentally retarded patients.

For sensitivity analysis, each input parameter changed ±1SD, and its effect on ICER, incremental costs, and incremental effects were calculated and shown by tornado plots.

Results

The estimated cost and effectiveness parameters for each intervention are shown in Table 1.

For different scenarios and with different screening TSH cut-off points, ICERs were similar. With a cut-off of 5 mU/l, ICERs were estimated at -4.58 (95% CI: -4.95, -4.21) thousand US dollars/gained DALYs; for 10 mU/l, ICERs were estimated at -4.57 (95% CI: -4.94, -4.20); for 15 mU/l, ICERs were estimated at -4.50 (95% CI: -4.87, -4.13); and for 20 mU/l, ICERs were estimated at -4.45 (95% CI: -4.82, -4.07) thousand US dollar per gained DALYs. The negative estimations have shown that in this model with the above-mentioned assumptions, execution of the screening program with different TSH thresholds would result in cost-saving and effectiveness. Acceptability curves also show that saved costs/gained DALY in different scenarios are close (Figure 2).

The distribution of our patients’ screenings, with regards to TSH was: 26.7% (5 – 9.9 mU/l), 36.7% (10 – 14.9 mU/l), 13.3% (15 – 19.9 mU/l), and 23.3% (≥ 20 mU/l). Therefore, we expect more gained DALYs and more cost savings with lower TSH cut-off points (Figure 3). In our theoretical cohort of 10,000 neonates, the estimated gained DALYs are calculated as: 316 ± 50 (95% CI: 220 – 415) for CH screening programs with a TSH threshold of 5 mU/l; 251 ± 40 (95% CI: 175 – 330) for 10 mU/l; 146 ± 23 (95% CI: 101 – 192) for 15 mU/l, and 113 ± 18 (95% CI: 79 – 149) for 20 mU/l. With such a difference in gained DALYs between each scenario, and the consequent difference in education and care of mentally retarded patients, it is evident that lower TSH thresholds result in greater gained DALYs and more significant savings in the cost of education and care of mentally retarded patients.

Sensitivity analysis by constructing tornado plots demonstrate the robustness of the model in which changing the input parameters ±1SD results in less than 0.5 SD in outputs (ICERs, incremental costs, and incremental effects). However, the greatest impact is pertinent to discount rate. Since the sensitivity results for all scenarios are approximately identical, only the tornado plots of the CH screening program with a TSH cut-off point of 5 mU/l have been presented in Figure 4.

Discussion

This study demonstrated that the CH screening with a threshold of 5 to 20 mU/l or more was economic. Increasing the cut-off point resulted in decreased recall rate of neonates, repetition of tests, and the psychological load and stress of families. However, the study showed that with elevating the TSH cut-off point screening, the number of life years without disability (gained DALYs) significantly decreased; the cost of education and care of patients with CH increased. This is a difficult situation for health decision makers because increasing the cut-off point screening leads to losing many patients and significant costs and resources. Thus, other countries act on lowering the cut-off point screening to decrease the numbers of false-negative cases.\textsuperscript{13,15,22–24}

Most similar studies quantified costs and benefits in monetary terms with no attempt to introduce a measure of health improvement into their analysis.\textsuperscript{13,15,22–24} We incorporated gained DALY as a health benefit measure from screening to quantify health effects. DALY is a known measure for comparing effects of different interventions for a disease or different diseases. Without any health measures, we could not differentiate between those interventions affecting health from those reducing the costs of the disease (i.e., cost of care and education). There was only one other study in the field of CH screening CEA that focused on saved DALY as a benefit.\textsuperscript{25} This study has indicated that to prevent each DALY in CH patients with the current screening program in Iran, approximately 28 dollars were utilized. It has also shown that each DALY costs about 624 dollars (accounting for education and care). The results of this research were not close to the findings obtained in our study. The disability weight of CH patients with mental retardation was regarded as equal to 0.35 for all ages, which was not in agreement with the Dutch burden of diseases study. Also, the researchers considered the cost of education and care only in daily public centers, which made all cost estimations use the lowest possible conditions.

According to the findings of the study, this model can measure the effect of different factors on effectiveness of the CH screening program. However, if there are more complete data in this model, yield results assist to decision making. For example, we have no exact information on these subjects if the screening model changes or if no preventive action is taken for pregnant women. Also, if the prevalence of disease varies, it affects the screening TSH distribution of neonates who have TSH more than 5 mU/l and/or it changes the ratio of transient to permanent cases. Clarifying these points will complete the disease model and obtain more functional data.

The estimation of disease cost should be reviewed exactly and indirect expenses should be considered and calculated. However, the high direct cost of this disease, the effectiveness of screening programs from one side, and the stability of the cost-effectiveness model against variation of cost on the other side demonstrate that we can emphasize the efficiency of the screening program in
the society. However, we should collect more information from patients who have been diagnosed outside of the screening program, because the assumption in which all neonates will become mentally retarded is unrealistic. According to the Geelhoed study patients are categorized into four groups: severe, middle, mild (borderline), and normal.14

This study shows that the current threshold of TSH at the screening program in terms of cost-effectiveness is the most suitable cut-off point. Still, more research is needed to reduce recall rate, the number of inappropriate repeat tests, and the psychological burden on families. It is evident that the results of this study are valuable for the current population. Also, it is necessary to update the results by changing the disease pattern as well as the components of the screening program, so that the most appropriate decision can be made by decision makers.

Acknowledgments

We would like to express our gratitude to Dr. Maziar Moradi (specialist in social medicine), Mrs. Sarrami, and Mrs. Bakhthiari (technicians in the neonatal screening laboratory) for their cooperation in this research. This study was financially supported by the Tehran University of Medical Sciences.

References