Nationwide Neonatal Screening for Congenital Adrenal Hyperplasia in Sweden
A 26-Year Longitudinal Prospective Population-Based Study

Sebastian Gidlöf, MD, PhD; Anna Wedell, MD, PhD; Claes Guthenberg, PhD; Ulrika von Döbeln, MD, PhD; Anna Nordenström, MD, PhD

**IMPORTANCE** Recent reports have questioned the rationale for neonatal screening for congenital adrenal hyperplasia (CAH) owing to low sensitivity in salt-wasting forms and a high rate of recall (ie, a positive finding resulting in a visit to a pediatrician and a second test) in preterm infants.

**OBJECTIVE** To determine the efficiency of the neonatal screening program for CAH in Sweden over time.

**DESIGN, SETTING, AND PARTICIPANTS** Longitudinal prospective population-based study in Sweden. We assessed neonatal screening for CAH from January 1, 1986, through December 31, 2011, when 2,737,932 infants (99.8%) underwent testing. The CYP21A2 genotype was investigated in 219 cases with true-positive findings (94.8%). We investigated the screening outcomes for 231 patients who had true-positive findings, 43 with late diagnosis, and 1,497 infants with false-positive findings.

**MAIN OUTCOMES AND MEASURES** Sensitivity of the screening for salt-wasting CAH. The most important secondary outcome measures were the positive predictive values and recall rates for full-term and preterm infants and sensitivity for milder forms of CAH.

**RESULTS** A total of 143 patients with salt-wasting CAH were identified; none were missed. The sensitivity was lower for milder forms of the disorder ($P = .04$), including 79.7% for simple virilizing forms and 32.4% for nonclassic forms. The positive predictive value was higher in full-term (25.1%) than preterm (1.4%) infants and correlated with gestational age ($r = 0.98; P < .001$). The recall rate in full-term infants (0.03%) was lower than that in preterm infants (0.57%) ($P < .001$). An analysis of previously reported results from other screening programs revealed that the sensitivity of the screening was negatively correlated with the duration of follow-up ($P = .03$).

**CONCLUSIONS AND RELEVANCE** Screening for CAH was highly effective in detecting the salt-wasting form and thereby reducing mortality. Additional late-onset cases of CAH were detected in childhood and adolescence, reducing the sensitivity for milder forms. The positive predictive value was high despite a low recall rate in full-term infants. Further improvements are necessary to increase the effectiveness of screening among preterm infants.
Congenital adrenal hyperplasia (CAH) affects 1/10,000 to 1/20,000 newborns. In more than 90% of the cases, CAH is caused by mutations in the CYP21A2 gene (OMIM 613815), resulting in 21α-hydroxylation deficiency. This mutation leads to deficiency of aldosterone and cortisol and accumulation of the precursors, 17α-hydroxyprogesterone and adrenal androgens. According to the severity of the 21α-hydroxylation deficiency, symptoms may vary from a life-threatening salt crisis with hyperkalemia and hyponatremia in the neonate to less dramatic signs of excessive androgen levels later during childhood or infertility in adults. As a result of fetal overproduction of androgens, girls are born with varying degrees of virilization of the external genitalia. The different clinical presentations are classified as salt-wasting, simple virilizing, and nonclassic CAH.

The results of the first screening program for CAH were published in 1977, and neonatal screening for the disease is currently performed in at least 30 countries. In Sweden, neonatal screening for CAH was started in 1986, with the 17α-hydroxyprogesterone level used as a marker of the disease. The cutoff level in full-term infants was gradually lowered from 50 to the present 60 nmol/L.

A patient was recalled after a positive result of screening. All children with a positive result of the first screening test were considered to be recalled. The patient visited a pediatrician and a second sample was obtained. Hence, a second filter paper sample was requested for all cases with a first positive result of screening. The second sample was usually analyzed about 1 week after the first one. From the laboratory perspective and based on clinical experience, these findings in full-term children with moderately elevated 17α-hydroxyprogesterone values who had convincingly decreased levels in the second sample (≥20%) were defined as false-positive if the clinical examination also excluded other clinical signs of CAH. This procedure has been used since the middle 1990s.

All pediatricians in Sweden were asked to report all cases of CAH to the screening laboratory, and all genetic investigations of CYP21A2 were performed at a single Swedish laboratory. A centralized registry of patients with CAH in this small country enabled a thorough follow-up of late-onset cases and those with false-negative findings.

Genetic Analysis and Classification of Disease Severity
The CYP21A2 mutation analysis was performed as described. Genotypes were divided into 5 groups according to the allele with the mildest mutation, as described previously. The salt-wasting CAH group consisted of the patients with the null and I2 splice genotypes. The simple virilizing group consisted of patients with the I172N or the P30L mutations as their mildest mutation. The nonclassic group consisted of patients with a clinical diagnosis and those with the V218L, the R341W, or the P453S genotype.

Statistical Analysis
Sensitivity was calculated as the proportion of cases with true-positive findings divided by the sum of those with true-positive and false-negative findings. We used the 2-tailed, unpaired t test for all analysis of temporal measurements, such as age at blood sampling and turnaround time at the laboratory. We used the χ² test to compare the following proportions: the differences in the rate of positive findings, PPV, and sensitivity between the sexes; the female to male ratio among cases detected in the screening and the background population; the difference in recall rates between preterm and full-term infants; and the proportions of cases with true- and false-positive findings with increasing 17α-hydroxyprogesterone values between the first and second test. We used the Spearman rank correlation test to analyze the correlations between the PPV and gestational age and between the 17α-hydroxyprogesterone level and CYP21A2 genotype. The Pearson product moment correlation test was used to correlate the

Methods
Screening
This study was approved by the Research Ethics Committee of Karolinska Institutet (No. 2010/1869-31/I). Informed consent was not required. Neonatal screening in Sweden is centralized to a single national laboratory. Until November 2007, blood samples were collected on Guthrie filter papers at 72 to 120 hours of age; after November 2007, samples were collected as soon as possible after 48 hours of age. Samples were sent by regular postal service to the laboratory.

Until 1991, 17α-hydroxyprogesterone was analyzed by radioimmunoassay and thereafter by a dissociation-enhanced, lanthanide fluoroimmunoassay (Delfia; Wallac Oy Corporation). We used gestational age–related cutoff plasma levels. The cutoff level in full-term infants was gradually lowered from 200 to 150 nmol/L in 1988 and to 75 nmol/L in 1991. With the introduction of a new antibody in 2005, the recall rate increased considerably and the cutoff level was adjusted to 100 nmol/L. The currently used antibody was introduced in 2008 and resulted in a lower cutoff level, which was adjusted stepwise from 50 to the present 60 nmol/L.

Until 2000, preterm infants (born before 37 full weeks of gestation) had a cutoff level of 200 nmol/L after ether extraction. Ether extraction was subsequently excluded because it delayed the result without considerably increasing sensitivity. The cutoff level for infants born before 35 full weeks of gestation was adjusted to 400 nmol/L, and neonates born at 35 to 36 full weeks of gestation were recalled if the screening value was 150 nmol/L or greater. With the introduction of the present antibody, the cutoff level was adjusted to 350 nmol/L before 35 weeks of gestation and to 100 nmol/L for 35 and 36 weeks.

A patient was recalled after a positive result of screening. All children with a positive result of the first screening test were considered to be recalled. The patient visited a pediatrician and a second sample was obtained. Hence, a second filter paper sample was requested for all cases with a first positive result of screening. The second sample was usually analyzed about 1 week after the first one. From the laboratory perspective and based on clinical experience, these findings in full-term children with moderately elevated 17α-hydroxyprogesterone values who had convincingly decreased levels in the second sample (≥20%) were defined as false-positive if the clinical examination also excluded other clinical signs of CAH. This procedure has been used since the middle 1990s.
reported sensitivity and years of follow-up in previous studies11-29 (Table 1).

In 78 cases, the 17α-hydroxyprogesterone level exceeded the standard curve (>600 nmol/L). For 42 of these cases, the exact value of 17α-hydroxyprogesterone was determined after dilution of the sample. However, for the remaining 36 cases, we did not determine the exact 17α-hydroxyprogesterone values. In the statistical analysis, these 36 cases were given the mean 17α-hydroxyprogesterone value for the diluted tests. The Mann-Whitney test based on rank rather than the mean was therefore used for statistical comparison. The 17α-hydroxyprogesterone values were expressed as median (95% CI). The age when the blood sample was obtained, time from sampling to arrival of the sample at the laboratory, time for completed analysis, and age at diagnosis are recorded. The results are given as mean (SD) for all cases with positive findings.

We used commercially available software (SPSS, version 20.0; SPSS, Inc) for all statistical analyses. Statistical significance was set at \( P < .05 \).

### Results

From January 1, 1986, through December 31, 2011, 2,742,944 infants were born alive in Sweden.30,31 Of these, 2,737,932 (99.8%) were included in the neonatal screening for CAH. Altogether, 1728 newborns were recalled owing to positive CAH findings for a total recall rate of 0.06%. Of these infants, 854 were preterm and 874 were full-term. The 17α-hydroxyprogesterone cutoff level was lowered stepwise to increase sensitivity; this process also increased the recall rate over time (Figure 1). The overall recall rate from 1986 to 2011 was significantly higher in preterm (0.57%) than full-term (0.03%) infants (\( P < .001 \)).

A total of 274 newborns who underwent screening in Sweden during the study period had CAH. In 231 newborns, the screening test result was positive, yielding an overall sensitivity of 84.3% for all forms of CAH. The overall sensitivity for boys (87.2%) did not differ significantly from that for girls (81.6%) (\( P = .29 \)). The specificity was 99.9%. A total of 12 pre-
term infants (5 boys and 7 girls; gestational age, 24-36 full weeks) were diagnosed as having CAH through the screening program.

The PPV was 25.1% for full-term infants. Despite gestational age–related cutoff levels, the PPV was low for preterm infants (1.4%) and correlated with gestational age ($r = 0.98; P < .001$) (Figure 2). The negative predictive value for full-term and preterm infants combined was close to 100%. We observed no difference in the female to male ratio of cases detected in the screening (115:116) compared with the general population ($133455:1409489$) ($P = .77$). Also, the female to male ratio in the entire population with CAH (141:133), in-
Genotype Groups and False-Positive Findings

<table>
<thead>
<tr>
<th>Genotype Group</th>
<th>Positive Finding</th>
<th>False-Positive Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>49 (n = 61)</td>
<td>12 (n = 12)</td>
</tr>
<tr>
<td>I2 Splice</td>
<td>82 (n = 82)</td>
<td>15 (n = 15)</td>
</tr>
<tr>
<td>I172N</td>
<td>49 (n = 49)</td>
<td>7 (n = 7)</td>
</tr>
<tr>
<td>P30L</td>
<td>6 (n = 6)</td>
<td>0 (n = 0)</td>
</tr>
<tr>
<td>P218L, R453S, and R341W</td>
<td>12 (n = 12)</td>
<td>1 (n = 1)</td>
</tr>
<tr>
<td>V281L</td>
<td>1 (n = 1)</td>
<td>0 (n = 0)</td>
</tr>
<tr>
<td>P453S, and R341W</td>
<td>12 (n = 12)</td>
<td>1 (n = 1)</td>
</tr>
<tr>
<td>R246C</td>
<td>2 (n = 2)</td>
<td>0 (n = 0)</td>
</tr>
</tbody>
</table>

Results are given for the 1728 infants with positive findings on the first screening. The null genotype group includes those with gene deletion or R356W, R443GGctc, Q318X, I172N + R356W, L307insT + Q318X, G291S, cluster E6 + V281L, R356P, R356+L307insT, Q318X + R356W, I172N + cluster E6 + V281L + L307insT, or R354H mutations. Error bars represent 95% CIs for medians. The salt-wasting CAH group consisted of the patients with the null and I2 splice genotypes. The simple virilizing group consisted of patients with the I172N or the P30L mutations as their mildest mutation. The nonclassic group consisted of patients with a clinical diagnosis and those with the V218L, R341W, or P453S genotype.

A total of 94.8% of the cases with true-positive findings underwent genotyping. In borderline cases with 17α-hydroxyprogesterone levels to 200 nmol/L, the 17α-hydroxyprogesterone levels increased between the first and second screening tests in 28 of 36 cases with true-positive findings compared with 30 of 467 cases with false-positive findings (P < .001).

Two infants with verified CAH died in the neonatal period despite having the CAH detected by screening. One was born at 25 full weeks of gestation and survived cardiac insufficiency and bradycardia during the first hours of life but died of septicemia at a few weeks of age. The other was born at 34 full weeks of gestation, developed lactic acidosis within the first week of life, and died at 2 weeks of age. In another 39 cases with elevated screening results, the infants died of unknown (to us) causes after the first screening sample with positive findings had been collected. A total of 19 of these infants were born before 28 full weeks of gestation, and only 13 were full term. The median 17α-hydroxyprogesterone levels for preterm (476 [95% CI, 420-630] nmol/L) and full-term (107 [77-173] nmol/L) infants who died were below the median 17α-hydroxyprogesterone level for infants with salt-wasting CAH as defined by the null and I2 splice genotype groups (792 [95% CI, 750-792] nmol/L) (P = .01 and P < .001, respectively).

The mean age at blood sampling in the cases with true-positive findings was 4.0 (2.4) days (decreased from 4.5 [2.5] to 2.4 [1.1] days with the introduction of the earlier sampling procedure [P < .001]. The mean administrative and posting...
time was 3.2 (1.7) days. The mean age of the child at the time of the final result in cases with true-positive findings was 8.7 (3.0) days. The mean age at blood sampling was significantly lower in girls (3.4 [2.7] days) than in boys (3.8 [1.3] days) ($P = .004$), in cases with salt-wasting CAH (null and I2 splice genotypes) (3.3 [1.4] days) compared with cases with milder forms (3.9 [1.2] days) ($P = .01$), and in full-term (3.4 [1.3] days) compared with preterm (4.5 [1.0] days) infants ($P = .001$). The turnaround time at the laboratory was significantly shorter in salt-wasting (1.8 [2.0] days) than in milder (2.4 [2.2] days) CAH ($P = .02$). The turnaround time at the laboratory decreased significantly throughout the study period ($P < .001$). From 1986 to 2011, blood samples for 33 tests with positive findings were obtained before 48 hours of age, and 15 of these results (all girls) were true-positive.

### Discussion

The screening program in Sweden successfully identified all infants with salt-wasting CAH. The sensitivity in the null and I2 splice genotype groups was high (99.3%), and no cases with salt-wasting signs were missed. The simple virilizing form had a sensitivity of 79.7%. The primary aim of the Swedish screening program is to prevent salt loss and to shorten the period of unclear sex. However, our 26-year follow-up, to our knowledge, the longest yet published, and the high proportion of patients with a known CYP21A2 genotype also enabled us to assess the extent to which patients with milder forms of the disease were identified by the screening.

The overall recall rate is low and the PPV is high in this study compared with those of other studies (Table 1), although these data varied depending on the different methods used over the years (Figure 1). The latest years show an improvement in the PPV for full-term and preterm infants owing to a more effective antibody used in the assay. We defined a recall as a positive result in the first sample, which is the most frequently used definition. Reports publishing higher PPVs and lower recall rates have defined a recall as a positive screening result in combination with findings of a clinical examination or a second test suggestive of CAH. The negative predictive value was close to 100%, reflecting the fact that CAH is a rare disease.

The 17α-hydroxyprogesterone levels used in the screening correlated with disease severity as determined by the CYP21A2 genotype, which confirms previously published results. However, because the ranges of 17α-hydroxyprogesterone levels within each genotype group were wide, we could not discriminate between different forms of CAH using the 17α-hydroxyprogesterone values alone. Cases with borderline true-positive findings were likely to have increasing 17α-hydroxyprogesterone levels in the second sample.

Although blood sampling occurred somewhat later in preterm infants (mean, 0.9 days older), the screening test seemed to be prioritized in the neonatology units. Blood sampling occurred earlier in girls compared with boys with true-positive results (mean age, 0.4 days), most likely as a result of suggestive clinical signs in virilized girls.

This study is unique in that it presents 26 years of newborn population-based screening for CAH covering 99.8% of all newborns in Sweden. Genotyping of CYP21A2 for confirmation and assessment of disease severity has made possible an accurate description of the sensitivity of the screening for different forms of CAH, including the mild cases. Furthermore, the national recommendations for care of CAH in Sweden include CYP21A2 mutation analysis. The screening and the genetic analyses are centralized to our unit. We therefore believe that, apart from the small risk for emigration before diagnosis, we report the true sensitivity of the Swedish screening program.

Owing to differences in health care systems between nations, the experiences of neonatal screening for CAH in Sweden are not directly transferable to all other countries. This study is limited by the fact that the screening program in Sweden is nationwide and no contemporary control group could be used. However, a follow-up study in Sweden investigated the clinical diagnosis vs diagnosis via screening alone. In the previous study, 73% of the boys and, surprisingly, 25% of the girls were identified via the screening. In addition, in a recent study using a historic control group, neonatal screening was effective in reducing mortality among boys and girls with CAH in Sweden.

In all screening populations, deaths before diagnosis, irrespective of the cause, are inevitable. As a result of physical stress, 17α-hydroxyprogesterone levels may increase in life-threatening conditions, such as septicemia or respiratory failure. Hence, severe illness apart from CAH may lead to a false-positive screening result. In the present study, the 17α-hydroxyprogesterone values were lower in infants with positive screening results who died than the levels detected in infants with the potentially life-threatening salt-wasting form of CAH. This finding suggests that the increased levels
of 17α-hydroxyprogesterone were due to nonadrenal disease and that death due to salt-wasting CAH is less likely.

The period of unclear sex was not assessed in this investigation. However, the previous follow-up study32 showed that the median period of wrong or unclear sex was shortened from 23 days to 3 days with the implementation of screening.

The rate of detection for salt-wasting CAH (100%) was high, whereas the overall sensitivity for all forms of the disorder (84.3%) was low compared with other reported data (Table 1). Most previously published studies have analyzed smaller series of patients, and the follow-up has been conducted during shorter periods (Table 1). Sensitivity is negatively correlated with years of follow-up when one combines the studies in Table 1 ($r = -0.52; P = .03$) because the true sensitivity of milder forms of CAH can only be evaluated when enough time has passed to allow diagnosis of late-onset cases. Thus, the somewhat lower reported sensitivity in the Swedish screening is partly owing to the long study period and a detailed follow-up of mild cases. In addition, the Swedish screening program is aimed primarily at detecting infants at risk for neonatal salt crisis, and the cutoff levels are therefore not set to detect all patients with nonclassic CAH. In fact, fewer than half of all identified patients with nonclassic disease were identified in the screening.

In contrast to previous findings,13 this study was not able to detect any difference in sensitivity between boys and girls. However, more boys had a positive screening result, possibly owing to testosterone production.

The efficiency of neonatal screening for CAH in preterm infants has been debated, and not all studies have shown beneficial effects of such screening. A recent publication4 concluded that because efficiency in preterm infants was low and the preterm neonates are carefully supervised on the neonatal ward, the screening for CAH in preterm infants should cease. Contrary to that suggestion, we consider screening to be justified because it does lead to earlier detection of patients, and the second sample can be obtained easily from admitted infants. In the light of our results, the screening laboratory finding can be informative about the low PPV in preterm infants, causing less worry to the families. A problem that remains to be solved, however, is the discrimination between false- and true-positive elevated 17α-hydroxyprogesterone values in this group. The preterm infants missed in the Swedish screening had more severe forms of CAH; they all belonged to the I172N genotype group, whereas the full-term infants with false-negative results predominantly belonged to the nonclassic V281I genotype group.

The results of this study covering 26 years of neonatal screening show that all patients with potentially lethal salt-wasting CAH can be detected with an acceptable recall rate. In combination with previous studies showing that screening leads to earlier detection32 and reduced mortality7 in salt-wasting CAH, this report further supports neonatal screening for CAH.

A high rate of false-positive results in preterm infants remains a problem. In a number of screening programs, the cutoff levels have been correlated with birth weight, gestational age, or both.33 Tandem mass spectroscopy and second-tier procedures may prove to be beneficial in preterm and full-term infants. Further studies addressing this issue are warranted.

Conclusions

The screening was highly effective in detecting salt-wasting CAH and thereby reducing mortality. Additional late-onset cases of CAH detected in childhood and adolescence reduced the sensitivity for milder forms. The PPV was high despite a low recall rate in full-term infants. Further improvements in the screening process are necessary to increase its effectiveness among preterm infants.

REFERENCES


