NEWBORN SCREENING FOR CLASSIC CONGENITAL ADRENAL HYPERPLASIA. UPDATE AND ASSESMENT OF REGIONAL PROGRAMS

Spanish full text

Introduction: the inclusion of an illness in a screening programme should be carried out in a rational and efficient way in the case of screenings for which sufficient scientific evidence on health benefits and cost-effectiveness studies is available. The effectiveness, safety and clinical efficiency of neonatal screening for classic congenital adrenal hyperplasia (CAH) was evaluated in 2013 and reported uncertainty in the fulfilment of 9 of the 18 essential requirements for implementation set out in the Spanish Ministry of Health's "Framework Document on Population Screening". In addition, the economic evaluation carried out in 2014 shown that their incorporation into the Spanish Newborn Screening Programme could be cost-effective, although with uncertainty in the decision. This HTA report is carried out as part of the working plan of the Network of Health Technology Assessment Agencies (RedETS) to analyse new scientific evidence on neonatal screening of CAH.

Objectives: to update the evidence on the clinical effectiveness of neonatal screening for CAH due to 21-hydroxylase deficiency; to reassess the degree of compliance with national requirements for implementing screening programmes agreed in the "Framework Document on Population Screening" to serve as a basis for decision making on their inclusion in the neonatal screening programme in the common basic service portfolio of the Spanish National Health System.

Methods: a systematic review of the scientific literature. A previous HTA report carried out in 2013 by Avalia-t was used as a starting point and the different search strategies were updated, taking into account the time limit used (2014 - present). The searches were carried out in the main biomedical databases: MedLine, Embase, Cochrane Library, HTA, DARE, INAHTA, CADTH, GIN, and Tripdatabse, among others, and the searches were regularly updated. The process was completed with a manual review of the bibliography of the articles included, and with a general internet search on the official websites of screening programmes, organisations and/or scientific societies. The titles and abstracts of the articles resulting from the search were reviewed by two researchers independently and blindly according to previously established selection criteria and resolving any discrepancies by consensus. The analysis also included data from the Spanish autonomous regions that include classic CAH screening in their additional screening programmes. This information was provided by the Screening Programmes Unit of the Spanish Ministry of Health. In order to incorporate the patients' perspective and to determine the degree of acceptability, a data collection survey was designed that included specific aspects aimed at understanding the impact of the disease on the patient, family members and caregivers, and the experiences and expectations related to the screening process.

Results and discussion: the systematic search retrieved 426 references, of which 29 were included because they met the previously defined selection criteria: 13 referring to screening programmes and 16 referring to disease characteristics. Through the manual review of the literature, the general internet search and regular updates up to April 2020, two additional studies, two economic evaluation

reports and a screening programme were found, making a total of 32 studies included. Through the manual search on official government websites, the results of 11 screening programmes were located, of which 5 were included, making a total of 18 studies and screening programmes included. All selected studies and screening programmes were population-based cohort studies with prospective data collection, six of which were carried out in Europe. The quality of the studies was evaluated according to the QUADAS-2 tool for diagnostic validity studies, generally obtaining a low risk of bias and a high applicability. For the section on epidemiology and disease characteristics, 2 clinical practice guidelines and 6 systematic reviews were included, 4 with meta-analysis, all of high quality, evaluated with the AGREE II and AMSTAR-2 instruments, respectively. To assess efficiency, an economic evaluation report by the Evaluation Service of the Canary Islands Health System (SESCS) and a Canadian costeffectiveness study were included, and to assess social aspects, quality of life and the experiences and perceptions of patients, families and caregivers, two systematic reviews, an integrative review and a descriptive study were included. By disseminating the survey to patient associations and scientific societies, the direct participation of one patient representative and 4 family members/caregivers was achieved.

Characteristics of classic CAH

CAH comprises a group of hereditary endocrine disorders characterised by adrenal insufficiency and varying degrees of hyper- or hypo-androgenic manifestations. 21-hydroxylase (21-OH) deficiency is the most common form, accounting for 90-95% of cases and leading to variable blockage of glucocorticoid (GC) and mineralocorticoid (MC) synthesis and excessive androgen production. The clinical variability of this 21-OH deficit is characterised by a very broad spectrum of symptoms whose severity depends on the functional impairment of the enzyme (total or partial) and of which two phenotypes are known: 1) the classic forms: these represent the more severe forms, beginning in the neonatal period and subdivided according to the degree of aldosterone deficiency into salt-wasting (SW) and simple virilizing (SV) forms and 2) the non-classical forms: these represent the moderatemild forms, manifesting themselves during late childhood, adolescence or adulthood with varying degrees of excess androgens and which at times are asymptomatic. In the SW form, there is a deficit of cortisol and adosterone that manifests itself in salt loss crises in the first weeks of life, which without treatment, develops into dehydration, metabolic acidosis and hypoglycaemia, coma and/or death. The SV form usually does not present salt loss crises, and newborns may present external genital alterations at birth that allow clinical diagnosis. The associated morbidity impacts on all stages of life, especially in childhood, and is associated with physical and developmental alterations that limit daily activities and generate psychosocial sequelae that significantly impact the quality of life; it often results in some form of psychiatric illness, behavioural and social adjustment problems; irreversible cognitive sequelae such as intellectual disability or problems with gender identity due to incorrect sex assignment at birth. Treatment consists of replacing hormone deficiencies in order to avoid saline crises and prevent lethal complications in the first weeks of life, as well as suppressing androgen synthesis to avoid virilisation, control the signs of androgenism and improve the consequences in adult life. For this purpose, supra-physiological doses of GC and MC are administered, until a balance between hyperandrogenism and hypercorticolism is achieved. However, chronic use of GC coupled with the complexity of regimen choice and dose adjustment increases the risk of adverse effects during

growth and adulthood associated with over- or under-treatment, which can affect development, cognitive and reproductive function, or the cardiovascular system, among others. In terms of prognosis, it is indicated that with adequate treatment patients can have a life expectancy similar to that of the general population, and that hormone replacement treatment administered at an early stage helps prevent salt loss and related complications before the onset of symptoms.

Characteristics of classic CAH screening programmes

The latency period of classic CAH is short, requiring a rapid response time especially in the SW forms that can lead to the death of the infant in the first weeks of life. Screening is performed by quantifying the level of 17-OHP in a blood sample (main biomarker) with time-resolved fluoroimmunoassay (TRFIA)

techniques, with the DELFIAR method being the most widely used. However, both 17-OHP and the other biomarkers used are also indicative of non-classical CAH, 11β-hydroxylase (11β-OHD) deficiency or other steroidogenesis deficiencies, so it is important to make a differential diagnosis by determining the steroid hormone profile. The test performance is limited and has low specificity for some 17-OHP antibodies, leading to cross-reactions with other adrenal steroids especially in premature infants. Causes of false positive results (FP) include prematurity and low birth weight, as well as concurrent disease, stress or steroid administration. In order to improve the test performance, a more specific antibody was developed for 17-OHP and it is recommended that a two-tier screening protocol is used: as a first tier, FIA is performed by stratifying cut-off thresholds according to gestational age and/or

birth weight, and as a second tier, liquid chromatography coupled to tandem masses (LC-MS/MS) is used to determine the steroid hormone profile. For the screening programmes evaluated in this review, 1228 cases of classical CAH (SW + SV) were detected, with a detection rate of 6.46 cases per 100,000 newborns. 1.2% of the cases were detected in Spain (5.48/ 100,000 newborns). According to the detection method, more cases were detected through screening than clinically (58% vs 42%) and mostly boys (85% vs 28%), while the proportion of girls detected clinically was higher (72% vs 15%). Taking into account the health status at the time of diagnosis, between 47% and 69% of patients detected by screening were asymptomatic, rising to 75% in the Spanish context. 79% (15/19) of the programmes used a one-stage screening protocol with FIA as the analytical method (DELFIAR,

AutoDELFIA, GSP) and 21% (4/19) a two-stage protocol, with FIA as the first level routine test and LC-MS/MS as the second level. Nationally, all of the autonomous regions used a one-step screening protocol, with FIA and mostly DELFIAR. In terms of screening algorithms and positivity thresholds, a high level of heterogeneity was observed, both nationally and internationally, which had a significant impact on the results. With identity of the protocol used, the sensitivity and specificity reported by the programmes were good. Sensitivity ranged between 94% - 100%, and specificity had values close to 100%. Programmes that used a one-stage protocol had a 5.5 times higher percentage of recalls than programmes that used a two-stage protocol. This implies a significant increase in the workload and the need for further testing, with the associated economic cost and emotional stress for families. One-stage screening programmes had a high number of FPs and twice as many as two-stage screening programmes. Concurrently, all studies showed worse results among preterm and/or low birth weight newborns compared to full term, estimating a difference between 2 and 10 times higher. The FP detected at national level was more than 4 times higher than international studies. The positive

predictive value (PPV) was low and close to 3% in one-stage screening, which increased to 5% in two-stage screening. For the set of programmes analysed, 26 false negative results (FN) were recorded, of which 65% were for SV forms and 27% for SW forms. Causes of FN include mother treatment with GC during pregnancy, administration of steroids due to prematurity or the presence of mild-moderate mutations. Based on the available data, the proportion of patients potentially benefiting from screening is estimated at 58% with important differences according to gender. By clinical suspicion, 42% of cases were identified including 30% of girls and 12% of boys, and by screening 49% of boys and 9% of girls were identified. Taking into account the health status at the time of diagnosis, between 47% and 69% of patients detected by screening were asymptomatic, rising to 75% in the Spanish context. Those most benefited by screening would be male newborns, as they usually do not present evident signs that facilitate its detection before the onset of an adrenal crisis, and they are more likely to delay diagnosis and to die from a potentially avoidable salt crisis. Screening would also make it possible to avoid incorrect sex-allocation in girls born with highly virilized genitalia or to reduce the time for correcting an incorrect allocation, thereby reducing the physical and psychological sequelae resulting from it, and ultimately reducing the time spent in hospital.

Conclusions: The evidence for the effectiveness of neonatal screening programmes of classic CAH assessed in this review is of moderate quality, and is based on population-based cohort studies with prospective data collection, systematic reviews with meta-analyses and clinical practice guidelines.

In classic CAH screening there is considerable heterogeneity in the screening algorithms in terms of cut-off stratification, recall criteria, reanalysis and/or referral to diagnostic confirmation, as well as in the protocols used (one or two-stage).

The test performance of single-stage screening programmes is limited, but is improved by adjusting positivity thresholds according to gestational age and/or birth weight and using the LC-MS/MS as a second-tier. The two-stage protocol reduces the false positive results and improves the positive predictive value. – It is estimated that in the absence of screening, about half of the cases may go undiagnosed or experience a diagnostic delay, especially in male newborns.

The updating of the evidence solves the existing doubts on 6 of the 9 essential requirements of implementation with uncertainty in the previous evaluation. Among these, the effectiveness of early treatment in improving the prognosis of the disease and the benefit in reducing morbidity stand out. However, the new evidence does not clarify the previous uncertainties about the benefit in terms of mortality, the validity of the test or the achievement of the expected benefit of the programme due to the short latency period requiring a rapid response time (especially in salt-wasting forms).

It is necessary to agree on an appropriate and specific screening algorithm and protocol that maximises the test performance and improves the comparability of the results, as well as to have a multidisciplinary clinical follow-up unit specialised in this illness.

It is recommended to establish specific indicators on health outcomes, and to define optimal and acceptable levels of some process indicators, such as the percentage of false positives. All this

information will help to measure the achievement of the established quality objectives and the decision-making process.