UK National Screening Committee

Screening for neuroblastoma

External review against programme appraisal criteria for the UK National Screening Committee (UK NSC)

Version: 1

Bazian Ltd.

February 2015

The UK NSC advises Ministers and the NHS in all four UK countries about all aspects of screening policy. Its policies are reviewed on a 3 yearly cycle. Current policies can be found in the policy database at http://www.screening.nhs.uk/policies and the policy review process is described in detail at http://www.screening.nhs.uk/policyreview

Template v1.2, June 2010

Abbreviations List

CI	Confidence Interval
COG	Children's Oncology Group
CT	Computerised tomography
EFS	Event-free survival
FDG PET	fluorine 18 fluorodeoxyglucose positron emission tomography
FPR	False positive rate
GN	Ganglioneuroma
GNB	Ganglioneuroblastoma
GPOH	German Pediatric Oncology and Hematology Group
HPLC	High performance liquid chromatography
HR	Hazard ratio
HTA	Health technology assessment
HVA	Homovanillic acid
INSS	International Neuroblastoma Staging System
INRG	International Neuroblastoma Risk Group
INRGSS	International Neuroblastoma Risk Group Staging System
IR	Incidence rate
IRR	Incidence rate ratio
JANB	Japanese Advanced Neuroblastoma Study Group
JINCS	Japanese Infantile Neuroblastoma Co-operative Study Group
MIBG	metaiodobenzylguanidine
MR	Mortality rate
MRI	Magnetic resonance imaging
MRR	Mortality rate ratio
MYCN	Myelocytomatosis viral-related oncogene neuroblastoma derived
NB	Neuroblastoma
OS	Overall survival
PPV	Positive predictive value
QoL	Quality of life
RR	Relative risk
RCT	Randomised controlled trial
SIOPEN	International Society of Pediatric Oncology Europe Neuroblastoma Group
TLC	Thin layer chromatography
VMA	Vanillylmandelic acid

Plain English Summary

Neuroblastoma is one of the most common forms of cancer in children under 5 years of age affecting around 100 children a year in the UK. The cancer is a form of tumour (an abnormal growth of tissue) found in particular nerve cells of the body. Neuroblastoma tumours may sometimes reduce in size or disappear with and sometimes without treatment but can also continue to grow and may rapidly progress to life-threatening disease. Children who develop Neuroblastomas that disappear without treatment are unlikely to ever show signs of being ill and screening all children may be the only way these tumours would be identified.

Depending upon the severity of the cancer, treatment may involve a mixture of surgery, chemotherapy and radiotherapy. These treatments can prevent death but can also be harmful and so it is important they are only given to those people whose cancer would not have disappeared without treatment.

It has been suggested that screening children of a certain age for Neuroblastoma may be the best method of ensuring cancer that would progress to life threatening disease is identified and treated early. Earlier treatment for those cancers is thought to be the best way of reducing the number of children dying of Neuroblastoma. The current UK NSC recommendation following the last review in 2005 is not to screen for Neuroblastoma in children. This is because the review found it was not possible to identify people that would benefit from treatment without also identifying people whose cancer would have improved without the dangerous treatment.

This review searched evidence between 2005 to November 2014 and found the quality of evidence was not high enough to confidently answer key questions about the benefit of Neuroblastoma screening. As such, the review has described the highest quality of research available and found:

- No evidence of a screening test that would only detect cases of Neuroblastoma that needed treatment and
- No evidence that Neuroblastoma screening (at any age) would reduce the number of deaths from Neuroblastoma.

Recommendation

There is insufficient evidence for the UK NSC to reverse its current policy not to recommend universal screening for neuroblastoma in children.

Executive Summary

The current UK NSC recommendation following the last review in 2005 is not to screen for neuroblastoma in children. This review and the 2003 HTA highlighted key uncertainties in the optimal age of screening, the screening strategy (single or multi-stage), the lack of a prognostic markers and the poor methodological design of studies to date.

This review will search the literature between 2005 to November 2014 and concentrate on answering two key questions in relation to whether there is evidence of:

- A screening strategy for neuroblastoma that reduces mortality rates and
- A prognostic marker that can effectively stratify risk and identify progressive forms of neuroblastoma

The quality of evidence was insufficient to accurately answer these key questions. As such, the review has described the highest quality of evidence available to answer these key questions and found:

- There is no high quality evidence that neuroblastoma screening (at any age) reduces mortality from the disease.
- No studies sufficiently addressed the primary concern that neuroblastoma screening could lead to over-diagnosis of biologically favourable cases who would otherwise have regressed spontaneously without treatment had they not been screen-detected.
- While some retrospective studies suggest screening at a later age may prevent over-diagnosis, currently no high quality studies have assessed the effect of screening at 18 months, either compared to no screening, or to screening at six or 12 months.
- Some retrospective studies have proposed several other disease characteristics and markers that may have prognostic significance. However, no prospective studies have assessed the accuracy of these variables for distinguishing between those of favourable and unfavourable prognosis. Furthermore these variables are not currently assessed in the first stage of screening, and therefore would not avoid the need for potentially unnecessary and invasive diagnostic testing in biologically favourable cases.
- Additionally, no high quality studies have been published that have assessed the effect of single compared to multi-stage screening.

Recommendation

There is insufficient evidence for the UK NSC to reverse its current policy not to recommend universal screening for neuroblastoma in children.

Introduction

Neuroblastoma

Neuroblastoma is a cancer that develops from neural crest cells, and is one of the most common embryonal tumours.[1, 2] Around 90 to 95 cases are diagnosed in the UK each year, most in children under the age of five.[1, 3] Neuroblastoma is the most common cancer in the first year of life, and is estimated to account for around one fifth[1] to one quarter[4] of all cancers in this age group. The aetiology of the condition remains unknown. Familial clustering has been observed, but such cases are rare and believed to account for less than 5%.[1, 5] Some research has suggested that certain exposures may be more common in children with neuroblastoma, but there have been methodological limitations with epidemiological studies to date.[5]

There are three distinct clinical patterns to neuroblastoma: spontaneous regression; maturation to benign ganglioneuroma; and rapid progression to life-threatening disease.[5] Clinical presentation depends on tumour site and whether there has been metastatic spread. Peak age of incidence is reported to be around 16 to 24 months.[2, 6] Initial symptoms may be general including lethargy, loss of appetite and pain if there has been spread to bone. The adrenal glands are the most common site for neuroblastoma development (about 40%[5]), which can cause the typical abdominal symptoms of distention, discomfort, constipation or difficulty passing urine. Tumours can also sometimes develop at pelvic, thoracic and cervical sites.[1-3]

Around half of all neuroblastomas have metastatic spread at the time of diagnosis,[6] though around 75% of children diagnosed with an abdominal tumour above the age of one year are reported to have metastatic spread.[2] Neuroblastoma diagnosed in the first year of life generally has better prognosis (83% five-year survival) than tumours diagnosed between the ages of one and four years (43% five-year survival).[1] Increasing age, advanced tumour stage, including metastatic spread and dissemination to bone marrow, and amplification of the *MYCN* gene are all recognised hallmarks of aggressive disease with poorer prognosis.[1, 5, 7]

Diagnostic and staging investigations involve ultrasound (frequently the initial investigation) followed by MRI or CT to stage the primary tumour. Radionuclide imaging with metaiodobenzylguanidine (MIBG) scintigraphy is performed in all patients. FDG (fluorine 18 fluorodeoxyglucose) PET CT may sometimes be performed in MIBG-negative patients.[2] Cytology of bone marrow aspirates and histology of core biopsies are also recommended at diagnosis.[1, 7]

The International Neuroblastoma Staging System (INSS) describe four main stages of neuroblastoma. Stage 1 is localised tumour with complete gross excision at surgery (there may or may not be residual microscopic disease remaining); stage 2 is localised ipsilateral tumour which can't be completely removed by surgery (2A negative lymph nodes but incomplete gross excision; 2B positive lymph nodes, with or without complete gross excision); stage 3 where there is involvement on both sides of the body (by either tumour or lymph nodes); and stage 4 where there are distant metastases. Stage 4S is a unique category given to children diagnosed at less than one year of age, where cancer may have spread to skin, liver or bone marrow, but less than 10% of bone marrow cells are neuroblastoma.[3, 6] Stage 4S tumours have a good outlook and often regress without treatment, so may be treated with a "watch and wait" approach.

Treatment of stage 1 to 4 neuroblastoma is dependent on tumour stage and grade, and may involve surgery, chemotherapy and radiotherapy. Intensive myeloablative regimens with high dose chemotherapy are increasingly being used to treat children with high-risk disease. There is evidence that these regimens may improve event-free survival compared with conventional chemotherapy in high-risk groups, though as yet no evidence that they improve overall survival.[8]

Screening for neuroblastoma

Historically, neuroblastoma may have seemed to be an ideal candidate for screening on many levels. It is one of the most common cancers in the first year of life, and children presenting before one year of age are known to have a better prognosis than those diagnosed at an older age. Most cancers are also already present at birth and can be detected by a simple, cheap and acceptable one-off urinary screening test. Nappies can be blotted with filter paper and urine tested for the catecholamine metabolites vanillylmandelic acid (VMA) and homovanillic acid (HVA), which are excreted by between 75 and 90% of neuroblastomas.[4]

Nationwide screening for neuroblastoma using a one-off urinary test at age six months was introduced in Japan in 1985, followed by feasibility programmes conducted in other countries in the 1990s. Two of these programmes, one in Canada (Quebec)[9] and one in Germany,[10] were non-randomised controlled studies that prospectively compared the incidence and mortality of neuroblastoma in screened and non-screened areas of the country.

There is no established optimal age for screening. Six months, as was used in Japan, has been the most commonly used age for a one-off screen. The Quebec programme performed screening at both three weeks and six months,[9] and the German programme performed screening at 12 months.[10]

There is also no standard laboratory test or cut-off level for urinary catecholamine metabolites. VMA and HVA levels can be measured qualitatively using gas chromatography or thin layer chromatography (TLC), or quantitatively using high performance liquid chromatography (HPLC). When nationwide screening was first introduced in Japan, VMA was measured using TLC. This was replaced nationwide in 1990 by the reportedly more accurate measurement of both metabolites using HPLC.[11] The Quebec programme measured VMA and HVA using gas chromatography, the German programme reportedly used HPLC.[11] When quantitative measurements are taken, with normalisation to urinary creatinine concentrations, the National Cancer Institute suggests cut-off levels of 25µg/mg creatinine for VMA and 32µg/mg creatinine for HVA, or alternatively two standard deviations above the laboratory's age-specific mean.[4]

Studies are reported to have demonstrated very high test specificity (approaching 100%), but lower sensitivity (between 40 and 80%).[4] The Quebec programme is reported to have demonstrated a positive predictive value (PPV) of 52%.[4] The German programme found a PPV of only 8.5%, and a false positive rate of 108.8 per 100,000.[10] False positives need to be followed for a prolonged period of serial non-invasive testing before the cancer can finally be definitely excluded.[4]

Importantly, there has been no evidence that screening for neuroblastoma reduces mortality from disease. The German programme found that screening led to an overall increase in the incidence of neuroblastoma (cumulative incidence 14.2 cases per 100,000 screened children vs.

7.3 cases per 100,000 controls).[10] However, this was reported to be due to an increase in the detection of biologically favourable, early stage neuroblastomas that may have regressed without ever presenting clinically. There was no increase in the detection of children with advanced stage disease with poor prognosis. The incidence of stage 4 neuroblastoma was similar in both screened and control areas of Germany (respectively, 3.7 and 3.8 cases per 100,000 children), and there was no difference in overall neuroblastoma mortality (1.3 deaths per 100,000 screened children, and 1.2 per 100,000 controls).[10] The German programme concluded that screening led to the diagnosis of an extra 7.0 cases per 100,000 children during the second year of life (excess cases not explained by earlier detection through screening).[10] Such over-diagnosis of biologically favourable cases can cause unnecessary diagnostic procedures and treatment, with the potential for attendant psychological and physical harms (e.g. treatment complications).

The Quebec programme similarly found no evidence that screening improved mortality. In Quebec the cumulative mortality from neuroblastoma was 4.78 per 100,000 children, which was similar and not significantly different from control groups in Ontario, Minnesota, Florida, and the Greater Delaware Valley.[9] The standardised mortality ratio for the Quebec screened cohort was 1.39 (95% confidence interval 0.85 to 2.30) compared to the rest of Canada.[9] Therefore screening did not reduce mortality from neuroblastoma in Quebec, which was in fact higher than that of some control groups. The small number of neuroblastoma deaths overall do mean though that some caution should be applied when interpreting these mortality figures. As the authors of the Quebec study suggest, there is still the possibility that screening may reduce neuroblastoma mortality. However, they consider it most likely that it has no effect. Furthermore the possibility cannot be ruled out that screening may cause harms though the unnecessary treatment of cases that would otherwise have run a benign course.

Following the results of the German and Canadian studies, mass screening in Japan was discontinued in 2003 on condition that incidence rates and mortality from neuroblastoma continued to be evaluated.[12]

The main challenge for the future of neuroblastoma screening is being able to have a reliable way of differentiating between children whose disease would progress and who therefore require treatment, from those whose disease would resolve spontaneously without treatment. Issues to be considered include the optimal age for screening and optimal screening strategy, and the identification of markers that can reliably distinguish between disease that will regress and progress.

Basis for current recommendation

The most recent 2005 UK NSC evidence review was based on an update search of literature from 2000 to June 2005, following the 2003 Health Technology Assessment (HTA), “A systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma”.[13] The 2005 UK NSC literature search found no robust new evidence that addressed the primary concern from the 2003 HTA regarding the clinical and cost-effectiveness of screening for neuroblastoma.

Several key uncertainties were highlighted by the 2003 HTA and the 2005 NSC evidence review:

- the optimal age of screening

- the screening strategy (single or multi-stage)
- lack of a prognostic markers
- the poor methodological design of studies to date (bias and lack of RCTs)

Following the 2005 review, the UK National Screening Committee concluded that screening for neuroblastoma should not be recommended because there was no evidence of its effectiveness.

Current update review

The current review considers whether the volume and direction of the evidence produced since the 2005 external review indicates that the previous recommendation should be reconsidered. Three main criteria will be considered, with particular focus given to areas the 2005 review identified as uncertain, or supported by insufficient evidence.

The main criteria and key questions reviewed are summarised in Table 1 below.

Table 1. Key questions for current neuroblastoma update review

Criterion	Key Questions (KQ)	# KQ Studies Included
2 - The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, disease marker, latent period or early symptomatic stage	Is there evidence of prognostic markers that can stratify risk and reliably distinguish between 'regressors' (those who may never present clinically if a screening programme hadn't identified them) and 'progressors' (those who would present clinically and progress through the stages of severity)?	3
13 - There should be evidence from high quality Randomised Controlled Trials that the screening programme is effective in reducing mortality or morbidity.	Does screening for neuroblastoma reduce mortality rates? Does focussing on detected cases at 18 months add any additional benefit (in terms of morbidity, QoL, detection/ FPR rate and cost) over screening at 12 months or 6 months? Is there evidence that either single or multi-stage screening is more beneficial than the other?	2

A systematic literature search of studies published between 2005 and 21/11/14 yielded 1064 references addressing neuroblastoma. Of these, 220 were assessed as being potentially relevant

to the key questions outlined in Table 1. These studies were further filtered at title and abstract level, and 54 were selected for appraisal at full text. Randomised controlled studies addressing each question were prioritised, followed by prospective non-randomised controlled studies, and finally retrospective studies. Each section below provides additional information on the evidence selection process for the given criterion.

Appraisal against UK NSC Criteria

These criteria are available online at http://www.screening.nhs.uk/criteria.

The following criterion is applicable to the first UKNSC key question:

2. The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, disease marker, latent period or early symptomatic stage.

Description of the previous UKNSC evidence review conclusion

The UK NSC concluded that screening for neuroblastoma should not be recommended because there was no evidence of its effectiveness. There is no reliable way of differentiating between children detected through screening who would be likely to progress and so would require treatment, from those who may never have presented clinically had they not been detected by screening.

Current UKNSC key question

The key question addressed by this review is:

- Is there evidence of prognostic markers that can stratify risk and reliably distinguish between 'regressors' (those who may never present clinically if a screening programme hadn't identified them) and 'progressors' (those who would present clinically and progress through the stages of severity)?

Description of the evidence

Overall 197 studies were identified as potentially relevant to this question following first-pass title and abstract sifting. Thirty-eight of these were further assessed at full text. Randomised controlled trials were to be prioritised, followed by non-randomised controlled studies (e.g. prospective cohorts with comparison to a control).

Since the 2005 NSC review no randomised or non-randomised controlled studies have been published that have compared the accuracy of one marker with another marker in identifying cases who will regress or progress. Additionally, no controlled studies were found that compared the accuracy of markers for predicting prognosis in terms of event-free survival or overall survival. This includes studies of markers that may be detected during initial screening (i.e. the urinary metabolites VMA and HVA), or markers that may be detected during subsequent diagnostic testing of screen-detected cases (e.g. imaging, serology, histology).

Only one study described as prospective was identified. Haber et al.[14] examined how *MRP1* gene expression (believed to have a role in mediating resistance to cytotoxic drugs) was associated with survival in a subset of 209 people from Pediatric Oncology Group (POG) Biology Study 9047 who had this information available from the time of diagnosis. However, this study was limited by the relatively small sample size. Therefore retrospective cohort study designs were also considered as the next level of evidence.

Since 2005 numerous retrospective cohorts have been published that have examined whether different disease characteristics or biological markers are associated with prognosis in neuroblastoma. Most of these studies have examined features that would be detected or considered during subsequent diagnostic testing (e.g. age, serology, histology, stage, genetic or tissue markers), rather than initial screening markers (i.e. urinary VMA and HVA). In terms of outcomes, studies have examined association with event-free survival or overall survival, or association with other recognised indicators of poor prognosis (e.g. advanced INSS stage or *MYCN* gene amplification).

Most studies have also been conducted in cohorts of people with a clinical diagnosis of neuroblastoma (or tissue samples from such cases), rather than in screened populations. A few retrospective cohorts have included screen-detected cases, though the aim of such studies has not been to examine how screening or diagnostic markers are associated with prognosis in screened compared with non-screened participants.

The decision was made to focus this review on three large retrospective cohorts from the International Neuroblastoma Risk Group (INRG) project. These studies include the population of POG 9047 as well as other groups, and therefore have a much larger collective sample size. They also looked at multiple rather than single markers, identifying those with the most prognostic significance.

The work from the INRG group has formed to date the largest dataset to have comprehensively examined the disease characteristics and markers that have the most significant association with prognosis. The aim of these studies was for the findings to be used to develop of a new pre-treatment risk stratification system. This system aims to help define homogenous pre-treatment patient cohorts and facilitate the assessment and comparison of baseline risk in clinical trials conducted in different parts of the world.

Numerous other retrospective cohorts have examined the association between different potential prognostic markers and either survival outcomes, or other prognostic indicators (e.g. advanced INSS stage or *MYCN* amplification). After the exclusion of small studies (fewer than 100 people), 25 further studies were excluded on the basis that they were currently only exploratory and retrospective in nature.

Results

Age cut-off in risk stratification

The INRG was a task force of investigators with expertise in neuroblastoma from major paediatric cooperative groups worldwide, which was established in 2004 with the aim of developing a consensus approach to pre-treatment risk stratification. The paediatric groups were the Children's Oncology Group (COG, North America and Australia), the German Pediatric

Oncology and Hematology Group (GPOH), the Japanese Advanced Neuroblastoma Study Group (JANB), the Japanese Infantile Neuroblastoma Co-operative Study Group (JINCS), and International Society of Pediatric Oncology Europe Neuroblastoma Group (SIOPEN).

The initial INRG cohort reported by London et al.[15] (Appendix 1) aimed to determine the optimal age-cut off as a prognostic variable in neuroblastoma. The study included 3,666 people diagnosed with neuroblastoma (confirmed by central pathology review) who were enrolled in COG trials between 1986 and 2001. London et al.[15] divided the cohort into 27 diagnostic age groups, with a concentration of age cut-offs in the 12-24 month age bracket, and retrospectively assessed the association with 4-year event-free survival (EFS). In analyses adjusted for INSS stage and *MYCN* status, EFS significantly decreased with increasing age category. Four-year EFS was 83% at diagnostic age less than 365 days (12 months) with EFS of 45% for those above this age cut-off ($p<0.0001$). When the age cut-off was at 460 days (15.1 months) there was an EFS of 82% for the cohort below this age and an EFS of 42% for those above ($p<0.0001$). Using a cut-off of 573 days (18.8 months) EFS was 74% for those diagnosed below this age and 38% for those above ($p<0.0001$).

When age was plotted against other prognostic variables of INSS stage, *MYCN* status, ploidy and Shimada histopathology, unfavourable prognostic features were reported to be more common in children diagnosed after the age of 12 months. This suggested that an age cut-off of 12 months was too low for risk stratification. When looking at the relative risk (RR) of EFS, p values were smallest for all age cut-offs between 15 and 22 months ($p< 1x10^{-30}$) with minimum p value at 18.8 months (573 days). Age cut-off of 15.1 months (460 days) had the greatest RR (2.598). The 4-year EFS for those diagnosed between 12 and 15.1 months was 73%, and 69% between 12 and 18.8 months.

London et al.[15] considered that any age cut-off within a range of 15 to 20 months would be statistically valid for future risk stratification, with 460 days (15.1 months) considered optimal.

Further development of pre-treatment risk stratification system
Cohn et al.[16] (Appendix 2) of the INRG task force aimed to further develop a consensus approach for pre-treatment risk stratification. Cohn et al.[16] conducted statistical analyses of prognostic factors; the companion publication by Monclair et al.[17] specifically outlined the INRG staging system (INRGSS) according to the presence of image-defined risk factors.

Cohn et al.[16] gathered data on 35 prognostic factors (consolidated into 13 factors for analysis) for 8,800 participants of the COG, GPOH, JANB, JINCS and SIOPEN trials who were diagnosed with neuroblastoma (NB), ganglioneuroblastoma (GNB) or ganglioneuroma (GN) maturing between 1990 and 2002.

Cohn et al.[16] repeated the analysis by London et al.[15] to identify an optimal age cut-off, excluding the 3,666 COG participants. Their result supported the findings of London et al.[15] that the optimal age cut-off was between 15 and 19 months. The consensus of the INRG task force was to use an age cut-off of 18 months (547 days) for risk stratification (though 12 months was used as the age cut-off to stratify risk in people with metastases and diploid tumours but without *MYCN* amplification [see Table 2]).

Appendix 2 (Table i) presents the clinical and genetic characteristics of the 8,800 participants in the INRG cohort which showed a statistically significant association with 5-year EFS and overall survival (OS). Cohn et al.[16] carried out survival tree regression analyses to identify the most highly significant variable. This created a branch in the survival tree in which the prognostic significance of the remaining factors was tested. This process was repeated with the remaining variables until the sample size within a branch became too small or no further significant variables were found.

Within the full cohort of 8,800 the presence of metastases (INSS stage 4) was the most significant prognostic factor. Five-year EFS when there were no metastases (INSS 1, 2, 3 or 4S) was 83% compared with 35% for stage 4 ($p<0.0001$). Respective OS was 91% for non-stage 4 and 42% for stage 4. Within favourable non-stage 4, histological category was the next most significant prognostic indicator (5-year EFS for GN maturing or GNB intermixed 97%, vs. 83% for NB and GNB nodular). Within non-favourable stage 4, age was the most significant prognostic indicator (5-year EFS for age <18 months 63%, vs. 23% >18 months).

Survival tree regression analyses led to a final consensus on seven variables that were clinically relevant and significantly associated with prognosis:

- INRG stage (as presented in Monclair et al.[17])
- age
- histologic classification
- grade of tumour differentiation
- *MYCN* status
- 11q aberrations
- tumour cell ploidy

These variables were put into the INRG consensus classification system which produced 16 clinically different pre-treatment risk groups from very low to high risk, as presented below (Table 2.)

Table 2: INRG consensus pre-treatment classification schema[16]

INRG Stage	Age (months)	Histologic category	Grade of differentiation	*MYCN* status	11q aberration	Ploidy	Pretreatment risk group
L1/L2		GN maturing or GNB intermixed					A: Very low
L1		Any, except GN maturing or GNB intermixed		Not amplified			B: Very low
				Amplified			K: High
L2	<18 months	Any, except GN maturing or GNB intermixed		Not amplified	No		D: Low
					Yes		G:Intermediate
	≥18 months	GNB nodular or NB	Differentiating	Not amplified	No		E: Low
					Yes		H:Intermediate

			Poorly or undifferentiated	Not Amplified			H:Intermediate
				Amplified			N: High
M	<18 months			Not Amplified		Hyperdiploid	F: Low
	<12 months			Not Amplified		Diploid	I: Intermediate
	12 to <18			Not Amplified		Diploid	J: Intermediate
	<18 months			Amplified			O: High
	≥18 months						P: High
MS	<18 months			Not amplified	No		C: Very low
					Yes		Q: High
				Amplified			R: High

Cohn et al.[16] report "for illustrative purposes, the proportion of patients when arbitrary EFS cut points are applied to cluster rows of the INRG consensus stratification". This is interpreted as follows:

- 28.2% of all diagnosed cases would fall into a very low risk group, and over 85% of this group achieve 5-year EFS (i.e. would survive to 5 years without events)
- 26.8% of all cases would fall into a low risk group, and between 75 and 85% of this group survive to 5 years event-free
- 9.0% would fall into an intermediate risk group, and between 50 and 75% of this group survive to 5 years event-free
- 36.1% would fall into a high risk group, and less than 50% of this group survive to 5 years event-free.

The companion publication by Monclair et al.[17] (Appendix 3) reports the INRG staging system according to the presence of image-defined risk factors (IDRFs). The principle was developed by SIOPEN who previously evaluated these as "surgical risk factors" – radiological features at the time of diagnosis that made safe and complete surgical resection less likely.[18] Monclair et al.[17] retrospectively assessed the association of IDRFs with 5-year EFS in 661 patients from SIOPEN with INSS stage 1, 2 and 3 disease who had known information on IDRFs at the time of diagnosis.

Monclair et al.[17] found that one or more IDRF (INGRSS stage L2) were present in 21% of people with INSS stage 1 disease, 45% with INSS 2, and 94% with INSS 3. For 474 people in the cohort with outcome data available, IDRFs were significantly associated with 5-year EFS, which was 90% for INGRSS L1 vs. 78% for L2 (p=0.001). IDRFs were also associated with 5-year OS (L1 96% vs. L2 89%; p=0.0068).

Monclair et al.[17] followed a study by Simon et al.[19](Appendix 4) who had also retrospectively assessed the impact of IDRF on surgical outcomes and prognosis. This was in a German trial population of 366 people with localised neuroblastoma. This study similarly found that the presence of IDRF (i.e. INGRSS L2) was associated with poorer 3-year EFS (L1 86% vs. L2 75%; p=0.01), though there was no significant effect on 3-year OS. However, multivariate analysis considering age, INSS stage and treatment approach did not find IDRFs to be an independent risk factor for an event (though INSS stage was).

Since the 2009 publications on the INRG classification system by Cohn et al.[16] and Monclair et al.[17] there have been further analyses from the INRG project. Schleiermacher et al.[20] specifically analysed 505 of 8,800 people in the INRG database who did not have *MYCN* amplification and for whom a genomic type could be attributed. They found that in people without *MYCN* amplification, a segmental genomic profile (any segmental chromosome alteration at 1p, 11q or 17q) was associated with poorer 4-year EFS (53%) than having no segmental alterations (79%).

Moroz et al.[21] retrospectively reviewed 11,037 cases in the INRG database looking back to between 1974 and 2002 to see whether there has been any change over three decades (1974-89; 1990-96; and 1997-2002) in the prognostic influence of age.

Overall 3-year EFS in all neuroblastoma cases have increased over time from 46% in 1974-89, 63% in 1990-96, to 71% in 1997-2002. For those diagnosed between ages 13 and 18 months, 3-year EFS is reported to have increased from 42% in 1974-89 to 77% in 1997-2002. For a diagnostic age above 18 months the respective increase is reported to have been from 25% to 45%. The hazard ratio (HR) for an event is still increased for those diagnosed above the age of 12 months compared to those younger, but the size of the increase in risk has decreased over the three decades. In 1974-89 the HR for those diagnosed aged 13-18 months was 3.94 compared with those ≤12 months, which decreased to 1.63 in 1997-2002. Similarly the HR for those diagnosed above 18 months of age was 4.61 compared with those ≤12 months in 1974-89, which decreased to 3.94 in 1997-2002. These results suggest that the influence of age on prognosis may have reduced over time.

It is possible that improvements in prognosis over time are a reflection of improvements in treatment approach. The authors report selecting the three eras for study as they represented distinct changes in therapeutic strategy: during 1974-89 multi-agent chemotherapy regimens were introduced but surgery and radiotherapy remained the main modalities; during 1990-96 regimens were risk-based with more intensive therapy for those at highest risk; from 1997-2002 most high risk patients received stem cell transplant. The change in the influence of age on prognosis may be a reflection of a change to more intensive treatment approaches for those above the age of 12 months in recognition of their higher risk status.

However, despite its possibly declining influence, the authors still consider age to have retained prognostic significance, alongside metastases and *MYCN* status. Moroz et al.[21] supported the earlier suggestion by Cohn et al.[16] of diagnostic age of 18 months as a pragmatic cut-off for the purposes of risk stratification.

Evidence quality

The INRG project, combining expertise from major paediatric cooperative groups worldwide, forms to date the largest dataset to have comprehensively examined the disease characteristics and markers that have a statistically significant association with prognosis.

However, the INRG studies have not examined whether levels of the urinary metabolites used in screening (VMA and HVA) have prognostic significance. Aside from age itself, all other factors in the pre-treatment risk stratification system are disease characteristics or markers that would be detected during subsequent diagnostic testing. Therefore the proposed INRG stratification system would not reduce the need for further (sometimes invasive) testing of screen-detected

cases who may have favourable outlook, and the attendant potential physical and emotional harms of this.

The INRG studies have also assessed prognostic significance in terms of association with 4- or 5-year EFS or OS, rather than examining whether particular factor(s) were clearly associated with whether a person's disease regressed or progressed without treatment.

The INRG explicitly state that they do not recommend that treatment be assigned according to risk stratification system that they developed. Nor do they suggest that the INRGSS is used as a substitute for the INSS. Their aim is that the pre-treatment risk stratification system is used to identify homogenous pre-treatment patient cohorts and facilitate the comparison of future, risk-based clinical trials. However, as yet no studies have been published that have prospectively evaluated the prognostic value of the INRG stratification system, or the presence of IDRF as laid out in the INRGSS.

The studies by London et al[15], Cohn et al.[16] and Monclair et al.[17] are all retrospective, which increases the possibility of inaccurate estimates of the prognostic significance of different variables. Data on all disease characteristics and markers of interest, and on survival outcomes, may not have been available for all members of the cohorts. For example, Monclair et al.[17] included only those from SIOPEN who had information available on IDRFs at the time of diagnosis, only a further 72% of whom had survival information available.

Cohn et al.[16] conducted survival tree analyses, which aimed to sequentially identify the next variable most significantly associated with survival in successive patient subgroups. This leads to gradually fewer patient numbers in each subgroup, which may give an inaccurate estimation of the true prognostic significance of any given variable. For this reason certain markers could not be examined. For example, Cohn et al.[16] could not examine the prognostic significance of 17q gain (segmental gain on the long arm of chromosome 17) because data on this was available for less than 5% of the cohort. Therefore all variables with potential prognostic significance may not have been examined, or identified.

It is further difficult to quantify with certainty the independent risk attributable to any given variable in unadjusted analyses, and it is possible that other factors (e.g. treatment approach) are confounding the association of any individual variable upon survival outcomes.

Overall prospective studies are needed to examine the accuracy of disease markers alone or in combination as indicators of favourable or unfavourable prognosis. Ideally this would include markers which could be used in screening.

Summary: Criterion 2 not met

No randomised or non-randomised controlled trials have been published that have assessed the prognostic significance of any disease characteristic or marker, either those that may be detected during initial screening or subsequent diagnostic testing.

Large retrospective cohort studies of people with neuroblastoma from the INRG project have identified various disease characteristics and markers that have the most statistically significant association with prognosis. They include age, imaging findings, tumour histology and grade of differentiation, *MYCN* status, variations of ploidy and segmental chromosome alterations. These factors have been formulated into the INRG pre-treatment risk stratification system.

However, in the context of screening, all of these factors (aside from age) are disease characteristics that could only be identified through subsequent diagnostic testing of cases initially identified by the presence of the urinary metabolites VMA and HVA. Therefore the proposed stratification system would not reduce the potential harms of unnecessary and potentially harmful intervention in children whose disease may have regressed and never presented clinically had they not been screen-detected.

Even with further diagnostic testing to establish the disease markers that make up the INRG risk-stratification system, it would not be currently possible to base treatment decisions on this system. To date no prospective studies have examined the accuracy of this set of factors, alone or in combination, in terms of identifying those with favourable or unfavourable prognostic outlook. Within any risk group it could not be known with any level of certainty who would require therapeutic intervention (and what the best treatment would be), and who would regress without treatment.

There is no currently no disease marker(s) that could be used in a screening programme for neuroblastoma that could reliably distinguish those with progressive disease who would require treatment, from those with neuroblastoma of favourable prognosis that would regress without treatment and who can therefore be spared unnecessary diagnostic and therapeutic interventions.

13. There should be evidence from high quality Randomised Controlled Trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screened to make an "informed choice" (eg. Down's syndrome, cystic fibrosis carrier screening), there must be evidence from high quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.

Description of the previous UKNSC evidence review conclusion

The UK NSC concluded that screening for neuroblastoma should not be recommended because there was no evidence of its effectiveness.

Current UKNSC key question

The key questions addressed by this review are:

- Does screening for neuroblastoma reduce mortality rates?
- Does focussing on detected cases at 18 months add any additional benefit (in terms of morbidity, QoL, detection/ FPR rate and cost) over screening at 12 months or 6 months?
- Is there evidence that either single or multi-stage screening is more beneficial than the other?

Description of the evidence

Overall 23 studies were identified as potentially relevant to these questions following first-pass title and abstract sifting. Sixteen of these were further assessed at full text. Randomised controlled trials were prioritised, followed by non-randomised controlled studies (e.g. prospective studies with comparison of screening to a control population).

Since the 2005 NSC review no randomised or non-randomised controlled studies of neuroblastoma screening at any age (e.g. 6 months, 12 months or 18 months) have been published, either compared to screening at another age or compared to no screening. There have also been no published randomised or non-randomised controlled studies assessing the effectiveness of single or multi-stage screening.

In the absence of any prospective studies, retrospective cohort study designs were considered as the next level of evidence. Since 2005 two retrospective population-based cohort studies have been published. These studies are both from Japan and have retrospectively assessed the effects of the discontinued nationwide screening programme upon incidence and mortality from neuroblastoma. One of these studies compared children who participated in the screening programme with contemporary non-participants; the other compared children born before and after introduction of nationwide screening.

In the absence of any higher level evidence, both of these retrospective cohorts were included in this evidence review. Excluded studies included case series, one small case-control study (comparing children from different countries by screening exposure), and smaller retrospective cohorts from within the population covered by the Japanese studies.

Results

The two retrospective cohort studies covered the Japanese neuroblastoma screening programme, which introduced one-off screening at six months of age nationwide in 1985. This was discontinued in 2003 following evidence from the non-randomised controlled studies from Germany and Quebec, Canada, which found no effect of screening on mortality.

Hisashige [12] (Appendix 5) specifically assessed neuroblastoma screening by quantitative measurement of VMA and HVA using HPLC, which was introduced across most of Japan from 1990. The study covered 25 regions where HPLC screening was performed for children born between 1984 and 1997. They compared the incidence of and mortality from neuroblastoma after six months of age for children screened by HPLC (n=3,705,670) with non-screened children (n=603,900) from the same population (i.e. children from the same areas who did not participate in the screening programme).

Hisashige[12] found that screening increased the incidence of neuroblastoma from six months to five years of age (708 screened cases vs. 59 unscreened). Overall this is broadly equivalent to 19.1 cases per 100,000 among screened children compared with 9.8 per 100,000 among non-screened children, but this does not take into account person-years of follow up. When looking specifically by stage and age at diagnosis, screening significantly increased the incidence of early stage neuroblastoma (INSS 1, 2 or 4S) diagnosed in children aged 6-12 months (incidence 255.54 screened vs. 26.72 non-screened children per million person years). However, screening had no effect on the incidence of advanced stage neuroblastoma (INSS 3 or 4) in children aged 6-12 months.

For children above one year, screening had no effect on the incidence of early stage neuroblastoma in children up to four years, but it was associated with significantly decreased

incidence of advanced stage in this age group (incidence 5.57 screened vs. 13.94 non-screened children per million person years).

The overall mortality rate from neuroblastoma (defined as death from neuroblastoma or related treatment) above the age of six months was significantly reduced by HPLC screening: 2.46 per million person years in the screened group vs. 4.50 per million person years in the non-screened group (mortality rate ratio 0.547, 95% CI 0.306 to 0.976). When looking specifically at age of death, screening was associated with significantly reduced mortality from neuroblastoma in the one to three year age group.

Hiyama et al.[11] (Appendix 6) conducted a before-after study design comparing children born in 1980 to 1983 before nationwide screening was introduced (n=6,130,423); children born in 1986 to 1989 during nationwide qualitative screening using TLC (n=5,290,412); and children born in 1990 to 1998 during nationwide quantitative screening using HPLC (n=10,868,860). It compared the incidence of and mortality from neuroblastoma before the age of six years for children born in these three periods.

Screening was associated with significantly increased incidence of neuroblastoma diagnosed up to the age of six years compared with the pre-screening period. When excluding cases diagnosed before six months, cumulative incidence per 100,000 children was: 10 in the pre-screening period, 19.42 during qualitative screening, and 27.82 during quantitative screening.

When looking by stage of diagnosis, screening was associated with significantly increased cumulative incidence of favourable stage 4S in children aged six months to six years: pre-screening 0.21 per 100,000, qualitative screening 0.78, and quantitative screening 1.49 per 100,000. Screening was also associated with significantly increased cumulative incidence of stage 1 to 3: pre-screening 3.99, qualitative 13.28 and quantitative screening 21.76 per 100,000. Qualitative screening had no effect on the incidence of advanced stage 4 compared to pre-screening, but quantitative screening was associated with decreased cumulative incidence of stage 4 (5.21 before vs. 4.27 per 100,000 during screening).

When analysed by uptake of screening in the two periods, the overall increased incidence of neuroblastoma between six months and six years was only significant among children who participated in screening. For quantitative screening, the difference in incidence of all stages compared to pre-screening was only significant for children who participated in screening. For qualitative screening, there was no consistent pattern between screening uptake and effect on incidence.

The overall neuroblastoma mortality up to six years was significantly decreased for children born during both screening periods compared to pre-screening. The overall cumulative neuroblastoma mortality rate up to six years was 5.38 per 100,000 during pre-screening, 3.90 during qualitative, and 2.83 during quantitative screening (respective RRs compared to pre-screening 0.73 [95% CI 0.58 to 0.90] and 0.53 [95% CI 0.42 to 0.63]). When excluding deaths prior to six months, the cumulative mortality rate between six months and six years was also significantly reduced by screening: 5.01 during pre-screening, 3.57 during qualitative, and 2.56 during quantitative screening (respective RRs compared to pre-screening 0.71 [95% CI 0.56 to 0.89] and 0.51 [95% CI 0.42 to 0.63]). Looking at screening participation, the reduced mortality in both screening periods was only significant among screening participants.

Evidence quality

Both of these retrospective cohort studies from Japan found that neuroblastoma screening was associated with significantly reduced neuroblastoma mortality compared to their respective non-screened comparison cohorts. However, there are significant limitations to the quality of this evidence. The retrospective, non-randomised study designs may introduce several sources of bias.

Hisashige [12] has compared participants and non-participants of HPLC screening from the same regional areas. The previous studies from Germany[10] and Quebec [9] prospectively compared screened and non-screened control regions over the same time period. Comparison of children within the same region may remove some potential bias resulting from differences between regions (for example, environmental or population characteristics or healthcare services). However, selection bias between participants and non-participants of screening is a significant limitation of the Hisashige study. There may be differences between those who have and have not chosen to participate in screening which may be confounding the results (e.g. parental socioeconomic status, age, health and lifestyle behaviours). The study does not report characteristics of screened and non-screened participants, and have not carried out any adjustment for differences in participant characteristics.

Hiyama et al.[11] used a before-after study design comparing children born before nationwide screening with those born during the periods of TLC screening and HPLC screening. Changes in incidence and mortality over time could reflect advances in diagnosis, treatments and healthcare resources, rather than reflecting the sole effect of screening.

Though both studies took data from reliable national registries, the retrospective design may allow for the potential for recall bias and missing or variably-recorded information. In particular in Hisashige [12], information on neuroblastoma incidence and screened status of the cases came from variable sources (including parental interview to assess screening participation), which could introduce inaccuracy.

Though both of the studies assessed mortality, they assessed neuroblastoma mortality in different age groups, making it difficult to compare their results. Hisashige [12] reported neuroblastoma mortality per million person years for children in different age groups (6 months to 1 year; 1 to 3 years; and >4 years) who did and did not participate in screening. Hiyama et al.[11] reported comparative neuroblastoma mortality per 100,000 children between six months and six years of age during the different screening eras. Neither study provided any information on 5 year event-free survival, overall survival or mortality rates from the time of diagnosis in screened compared with non-screened participants.

Both of the studies also found a significantly increased overall incidence of neuroblastoma. However, the differences in design of the studies and their categorisation of age and disease stage make it difficult to directly compare the results of the two. Hisashige [12] found an increased incidence of early stage neuroblastoma between the ages of 6 and 12 months (INSS stage 1, 2 or 4S) in screened compared to contemporaneous non-screened participants. This may provide some further evidence that screening may increase detection of cases that may have spontaneously regressed had they not been picked up by screening. There was no effect on early stage diagnoses above the age of 12 months. Hiyama et al.[11] found incidence of INSS stage 1, 2, 3 and 4S neuroblastoma in children aged in the wider age category of six months to six years increased during the screening years compared to pre-screening.

Both studies did indicate that quantitative screening using HPLC may be associated with decreased incidence of advanced stage in older children: Hisashige [12] finding decreased

incidence of stage 3 or 4 in children aged one to four years, and Hiyama et al.[11] finding decreased incidence of stage 4 in children aged six months to six years. However, due to the limitations of the two study designs and potential for bias it is not possible to draw any firm conclusions on whether screening may influence the incidence of advanced stage neuroblastoma in older children.

Neither of these retrospective cohort studies is able to provide any information on several of the key questions addressed by this review. Both of the studies compared one-off screening at six months of age with no screening (either before screening was implemented or among screening non-participants). Though it has been considered that screening at the later age of 18 months may reduce the detection of biologically favourable cases, no studies have yet been published comparing screening at 18 months with either no screening, or screening at younger ages.

The evidence from these two retrospective cohort studies does not inform on the optimal screening strategy. Hisashige [12] specifically examined neuroblastoma screening using quantitative measurement of catecholamine metabolites using HPLC compared to not screening. Hiyama et al.[11] separately compared qualitative measurement using TLC and quantitative measurement using HPLC with not screening. Neither study can reliably inform on the comparative performance of these two measurements; nor can the two studies provide any evidence on the value of single-stage compared with multi-stage screening.

The studies also provide no further information on outcomes other than incidence and overall mortality including five year disease- or event-free survival, morbidity and quality of life, or cost effectiveness.

Summary: Criterion 13 not met.

There is no evidence from randomised controlled trials that neuroblastoma screening (at any age) reduces mortality from the disease.

Since the 2005 NSC review, only two Japanese retrospective cohort studies have been published.[11, 12] These studies have compared the incidence and mortality from neuroblastoma in children screened at six months with non-screened children (either contemporaries who did not participate in the screening programme, or children born before nationwide screening was introduced). Though both of these studies found a reduction in neuroblastoma mortality in screened compared with non-screened controls, there are significant limitations to the quality of evidence provided by these retrospective cohorts. This includes significant potential for selection bias in one study, and confounding due to use of non-contemporaneous controls in the other.

Both of the studies also generally observed an increased incidence of early stage neuroblastoma, which may include cases that would have regressed spontaneously had they not been detected by screening.

Though 18 months has been suggested as a possible screening age to minimise over-diagnosis of tumours with a favourable prognosis, no randomised or non-randomised controlled studies have to date been published assessing the effect of screening at 18 months, either compared to no screening, or to screening at six or twelve months.

No randomised or non-randomised controlled studies have been published that have assessed the effect of single compared to multi-stage screening.

Conclusions

Implications for policy

This report assesses newborn screening for neuroblastoma against select UK National Screening Committee (UK NSC) criteria for appraising the viability, effectiveness and appropriateness of a screening programme. This topic was last assessed by an external evidence review in 2005, which concluded that screening for neuroblastoma should not be offered because there was no evidence of its effectiveness.

One-off neuroblastoma screening at the age of six months was previously offered nationwide in Japan. This was discontinued in 2003 following evidence from non-randomised controlled studies conducted in Canada and Germany that screening did not reduce mortality from neuroblastoma. These studies also suggested that screening may lead to over-diagnosis of biologically favourable cases that would otherwise have regressed without treatment.

The 2005 review therefore identified several key uncertainties related to the optimal age of screening, the best screening strategy (single or multi-stage), lack of prognostic markers to distinguish regressors from progressors, and the poor methodological design of studies to date (bias and lack of RCTs).

This review assessed key questions to determine if evidence published since the last review resolves any of these uncertainties. The body of evidence published since 2005 does not alter the conclusions of the previous NSC review, and does not support overturning the previous recommendation not to screen for neuroblastoma in the UK.

A summary of key findings for the assessed criteria is below.

Epidemiology, natural history and clinical course

A key concern is the possibility that neuroblastoma screening could lead to over-diagnosis of biologically favourable cases who would otherwise have regressed spontaneously without treatment had they not been screen-detected. This review aimed to see whether there were reliable disease markers that could distinguish between cases who would regress and who would progress. No randomised or non-randomised controlled studies have been published that have compared the prognostic significance of disease characteristics or markers, either those that may be detected during initial screening or subsequent diagnostic testing.

Numerous mainly retrospective cohorts have explored whether different disease characteristics or biological markers are associated with EFS and OS in neuroblastoma. Of these retrospective studies, those of the International Neuroblastoma Risk Group (INRG) project have been assessed in detail in this review, as they currently form the largest dataset to have comprehensively examined the disease characteristics and markers that have the most significant association with prognosis. The INRG have identified seven prognostic factors that have been developed into a consensus pre-treatment classification system: age, presence of image-defined risk factors (IDRFs, as indicated by INRG stage), histological classification, grade of tumour differentiation, *MYCN* status, 11q aberrations and tumour cell ploidy.

In the context of screening, all of these factors, aside from age, would currently be assessed during subsequent diagnostic testing, and would not be identified through the first stage of neuroblastoma screening. Use of this system would not therefore reduce the unnecessary and potentially harmful intervention in children whose disease may have regressed had they not been screen-detected. Even with further diagnostic testing, the proposed INRG pre-treatment

classification system is not currently aimed at helping to make treatment decisions. Prospective studies have not yet determined the reliability of these potential prognostic variables in identifying those with favourable or unfavourable disease outlook, or what treatments are most appropriate for each proposed risk group.

Evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity

This review aimed to address whether new evidence has emerged since the last NSC review suggesting that neuroblastoma screening reduces mortality. Also whether increasing the screening age to 18 months (the INRG consensus cut-off for risk stratification) offers additional benefit, particularly in terms of primarily detecting those with unfavourable disease characteristics while reducing the diagnosis of biologically favourable cases.

Since 2005 no randomised or non-randomised controlled studies have been published that have assessed screening at any age, compared to screening at another age, or compared to no screening. Two population-based cohorts from Japan have been published that have retrospectively assessed the effects of the discontinued nationwide screening programme. These studies compared the incidence and mortality from neuroblastoma in screened and non-screened children (either contemporaries who did not participate in the screening programme, or children born before nationwide screening was introduced). Though both of these studies found a reduction in neuroblastoma mortality in screened compared with non-screened controls, there are significant limitations to the quality of evidence provided by these retrospective cohorts. This includes significant potential for selection bias and confounding.

Furthermore, though 18 months has been suggested as a possible screening age to maximise diagnosis of biologically unfavourable tumours, no prospective studies have yet assessed the effect of screening at 18 months, either compared to no screening, or to screening at six or twelve months. Neither have any prospective studies assessed the effect of single compared to multi-stage screening.

Overall the uncertainties raised by the last NSC review remain, and there is no robust evidence that neuroblastoma screening reduces mortality.

Implications for research

Given the lack of high quality evidence identified for each question; randomised controlled trials, or at minimum non-randomised controlled trials (e.g. prospective cohorts with comparison to a control group), are needed to resolve uncertainties regarding screening for neuroblastoma:

- Prospective studies are needed to assess the effect of neuroblastoma screening at 18 months of age, either compared to no screening or screening at alternative age (e.g. 6 or 12 months), on neuroblastoma mortality and morbidity (including that associated with false positives and over-diagnosis of biologically favourable cases).
- Prospective studies are needed to assess the accuracy of different disease characteristics and markers that have been proposed in retrospective studies as potential prognostic indicators, including those in the INRG pre-treatment risk stratification system. Future studies would also need to determine whether any of these markers alone or in combination were accurate and precise enough on which to base

treatment decisions – that is, to reliably distinguish between those children whose disease would regress and therefore not require any form of treatment from those with disease that would progress through the stages of severity if untreated.

- Prospective studies are needed to determine the optimal screening strategy, including methods of detection of urinary metabolites, and whether a multi-stage screening strategy offers any benefits over single-stage screening.

Methodology

The draft update report was prepared by Bazian Ltd., and then adapted in line with comments from the National Screening Committee.

Search strategy

Background search – systematic reviews, guidelines, health technology appraisals

Background searches were carried out in the Cochrane Database of Systematic Reviews, Health Technology Assessment database, the Database of Abstracts of Reviews of Effectiveness, NHS Evidence, the National Guidelines Clearinghouse, the Guidelines International Network, Google Scholar and Google Advanced search from 2005 to 21/11/14. Searches were limited to English language studies.

Medline search strategy

1 exp Mass Screening/ (105391)

2 screen*.ti,ab. (499387)

3 Mass-screen*.ti,ab. (4599)

4 or/1-3 (535521)

5 exp Neuroblastoma/ (25532)

6 (neuroblastoma* or ganglioneuroblastoma* or ganglioneuroma*).ti,ab. (30919)

7 or/5-6 (36993)

8 4 and 7 (1389)

Key question 1 – randomised controlled trials, controlled studies, cohort studies

Focused searches for KQ1 were carried out in Embase and Medline using filters developed by the Scottish Intercollegiate Guideline Network (SIGN), and in the Cochrane Central Register of Controlled Trials from 2005 to 21/11/14. Searches were limited to English language studies.

Medline search strategy

1 exp Mass Screening/ (105391)

2 screen*.ti,ab. (499387)

3 Mass-screen*.ti,ab. (4599)

4 or/1-3 (535521)

5 exp Neuroblastoma/ (25532)

6 (neuroblastoma* or ganglioneuroblastoma* or ganglioneuroma*).ti,ab. (30919)

7 or/5-6 (36993)

8 4 and 7 (1389)

Key question 2 Medline search – randomised controlled trials, controlled studies, cohort studies

Searches for KQ2 were carried out in Embase and Medline using filters developed by the Scottish Intercollegiate Guideline Network (SIGN), and in the Cochrane Central Register of Controlled Trials from 2005 to 21/11/14. Searches were limited to English language studies.

Medline search strategy

1 exp Mass Screening/ (105391)

2 screen*.ti,ab. (499387)

3 Mass-screen*.ti,ab. (4599)

4 exp Prognosis/ (1158701)

5 (prognosis or prognostic).ti,ab. (381608)

6 exp Diagnosis/ (6777441)

7 (diagnosis or diagnostic$1).ti,ab. (1447899)

8 ("follow up" or "follow-up").ti,ab. (654081)

9 Urinalysis/ (4913)

10 (Urinalysis or urine or urinary).ti,ab. (379210)

11 Serum.ti,ab. (815736)

12 ((tumour or tumor) and marker$1).ti,ab. (81779)

13 Homovanillic Acid/ (6073)

14 Homovanillic acid.ti,ab. (4814)

15 Vanilmandelic Acid/ (1843)

16 ("Vanilmandelic acid" or "vanillyl mandelic acid" or " vanillylmandelic acid").ti,ab. (1129)

17 Catecholamines/ (34926)

18 Catecholamine*.ti,ab. (55480)

19 ("MIBG scintiscan" or "MIBG scan").ti,ab. (189)

20 Dopamine/ (66334)

21 Dopamine.ti,ab. (108115)

22 Ferritins/ (16689)

23 Ferritin*.ti,ab. (21586)

24 Phosphopyruvate Hydratase/ (6835)

25 ("Neuron specific enolase" or "Neuron-specific enolase" or NSE).ti,ab. (7313)

26 L-Lactate Dehydrogenase/ (37950)

27 ("Lactate dehydrogenase" or LDH).ti,ab. (40481)

28 (((MYC-N or MYCN or N-myc) and (gene or oncogene or proto-oncogene)) adj3 amplification).ti,ab. (901)

29 ("Chromosome 1p" adj3 (deletion or "allelic loss" or "loss of heterozygosity")).ti,ab. (147)

30 ("chromosome 17q" adj3 gain).ti,ab. (16)

31 exp Ploidies/ (65256)

32 (Ploidy or "hyperdiploid" or "near-diploid" or "near-tetraploid").ti,ab. (11771)

33 Nerve Growth Factor/ (5792)

34 ("nerve growth factor" or trka or "trk expression").ti,ab. (17036)

35 or/1-34 (8832467)

36 exp Neuroblastoma/ (25532)

37 (neuroblastoma* or ganglioneuroblastoma* or ganglioneuroma*).ti,ab. (30919)

38 or/36-37 (36993)

39 35 and 38 (18664)

8 4 and 7 (1389)

Quality

Several factors were assessed to determine the quality of the identified evidence, including study design and methodology, risk of bias, directness and generalisability of the evidence. Factors that were determined to be pertinent to the body of evidence identified for each criterion are outlined in the results section as well as the comment section of the Appendix tables. The overall level of evidence was assessed by considering the quantity, quality and consistency of evidence across the body of studies for each criterion reviewed.

Appendices

Appendix number	1
Relevant criteria	2
Publication details	London et al. Evidence for an Age Cutoff Greater Than 365 Days for Neuroblastoma Risk Group Stratification in the Children's Oncology Group. Journal of Clinical Oncology. 2005; 23(27): 6459-6465.[15]
Study details	Retrospective cohort study Country: US. Setting: Participants of the Pediatric Oncology Group (POG) and Children's Cancer

	Group (CCG) trials from 1986 to 2001 (now reported to be part of the Children's Oncology Group [COG]).
Study objectives	To identify the statistically optimal age cut-off as a prognostic variable.
Inclusions/study population	3,666 participants of the POG and CCG trials. Eligibility criteria: diagnosis of neuroblastoma confirmed by central pathology review; age ≤21 years; written informed consent. All studies had collected data on age at time of diagnosis and disease stage, and tumour tissue was required or strongly encouraged. Median age of participants: 573 days; 64% (n=2,327) had age ≥365 days. Other characteristics: 44% (n=1,522) had INSS stage 4 disease, 18% (n=520) had *MYCN* gene amplification, 29% (n=483) had diploid tumours, and 43% (n=665) had tumours with unfavourable histopathology (not further defined).
Exclusions	None reported
Intervention/test and comparator	Not applicable
Prognostic variables examined	Diagnostic age: the cohort was divided into 20 age groups (19 age cut-offs), in addition to a concentration of 8 further age cut-offs tested in the bracket of 12-24 months.
Outcomes examined	4-year event-free survival (EFS) defined as relapse, disease progression, secondary malignancy, and death, whichever occurred first. Relative risk (RR) and p value were calculated for the 27 paired age groups using a multivariable Cox proportional hazards model, with adjustment for INSS stage and *MYCN* status. Median follow-up: 5.8 years.
Results/outcomes	<u>4 year EFS by diagnostic age cut-off</u> Significantly decreased with increasing age • 83% at age <365 days/12 months (n=1,339; 37%) vs. 45% at age ≥365 days/12 months (n=2,327; 63%): RR 2.333 (p<0.0001) • 82% at age <460 days/15.1 months (n=1,589; 43%) vs. 42% at age ≥460 days/15.1 months (n=2,077; 57%): RR 2.598 (p<0.0001) • 74% at age <573 days/18.8 months (n=1,833; 50%) vs. 38% at age ≥573 days /18.8 months (n=1,833; 50%): RR 2.494 (p<0.0001) <u>Determination of optimal age cut-off</u> • Age was plotted against the proportion of patients with known prognostic risk factors (INSS stage, *MYCN* status, ploidy and Shimada histopathology). In general before the curves crossed the cohort was

	dominated by favourable prognosis, and unfavourable prognosis after the cross. A crossing or confluence of curves occurred at ages > 365 days for *MYCN* status (12.3 months), stage (20 months), ploidy (14 months), and histopathology (17 months) • Looking at RR for EFS for each of 27 separate age-cut-offs, p values were smallest for all age cut-offs between 15 and 22 months ($p< 1x10^{-30}$) with minimum p value at 573 days (18.8 months). Age cut-off of 460 days (15.1 months) had greatest RR (2.598). There was increased risk of an event above 600 days (19.7 months). • 4 year EFS for 250 patients (7% of total cohort) with diagnostic age 365 to 460 days was 73%. Excluding higher risk patients with INSS stage 4 and *MYCN* amplification, 4 year EFS for the remaining 135 (5% of total cohort) was excellent at 92%. • 4 year EFS for 494 patients (13% of total cohort) with diagnostic age 365 to 573 days was 69%. Results of this study suggest that an age cut-off of 365 days is too low, as unfavourable risk factors dominate above 365 days. Any age cut-off within a range of 15 to 20 months was considered to be statistically valid for future risk stratification, with 460 days (15.1 months) considered optimal.
Comments	Retrospective data analysis increases the possibility of inaccurate risk estimates. There is the possibility of missing histological data, and that unmeasured factors are confounding analyses (e.g. treatment approach – POG and CCG participants were involved in therapeutic trials). Prospective studies assessing prognosis when risk is stratified by age are needed.

Appendix number	2
Relevant criteria	2
Publication details	Cohn et al. The International Neuroblastoma Risk Group (INRG) Classification System: An INRG Task Force Report. Journal of Clinical Oncology. 2009; 27(2): 289-297.[16]
Study details	Retrospective cohort study Country/Setting: International. A task force of investigators from the major paediatric cooperative groups worldwide: COG (North America and Australia), the German Pediatric Oncology and Hematology Group (GPOH), the Japanese Advanced Neuroblastoma Study Group (JANB), the Japanese Infantile Neuroblastoma Co-operative Study Group (JINCS), and International Society of Pediatric Oncology Europe Neuroblastoma

	Group (SIOPEN), between 1990 and 2002 (follow-up of data to 2004).
Study objectives	To develop a consensus approach for pre-treatment risk stratification of neuroblastoma based on statistical analyses of prognostic factors. The aim is to help facilitate the comparison of risk-based clinical trials conducted in different parts of the world by defining homogenous pre-treatment patient cohorts. The companion publication by Monclair et al.[17] outlines the new INRG staging system (INRGSS) according to the presence of image-defined risk factors. This publication reports the other INRG consensus classification schema.
Inclusions/study population	8,800 participants of the COG, GPOH, JANB, JINCS and SIOPEN trials. Eligibility criteria: confirmed diagnosis of neuroblastoma (NB), ganglioneuroblastoma (GNB) or ganglioneuroma (GN) maturing; age ≤21 years; diagnosis between 1990 and 2002; written informed consent; and information on the 35 potential risk factors as below. Participants in these trials were from North America (48% - COG), Europe (47% - SIOPEN and GPOH) and Japan (5% - JANB and JINCS).
Exclusions	None reported
Intervention/test and comparator	Not applicable
Prognostic variables examined	Information on 35 potential risk factors was collected: age*, INSS stage, Evans stage, Shimada classification, Shimada histologic category, Shimada grade, Shimada mitosis-karyorrhexis index (MKI), International Neuroblastoma Pathology Classification (INPC), INPC histologic category, INPC grade of tumour differentiation, INPC MKI, *MYCN* status, DNA ploidy (defined as DNA index ≤1.0 vs. >1.0), 11q loss of heterozygosity (LOH), 11q aberration, unbalanced 11q LOH, 1p LOH, 1p aberration, 17q gain, serum ferritin, serum lactate dehydrogenase (LDH), six primary tumour sites, and eight metastatic sites. These 35 potential factors were consolidated into 13 for analysis. They included the following dichotomous variables obtained from consolidation of some of the above factors: • INPC histological category (Shimada diagnosis, grade of differentiation, or MKI if unknown) • INSS staging criteria (Evans if unknown) • "11q aberration" as the combination of unbalanced 11q LOH and 11q aberrations • "1p aberration" as the combination of 1p LOH and 1p aberrations • Six primary tumour sites consolidated into adrenal vs. non-adrenal • Eight metastatic sites consolidated into metastases vs. no metastases

	Only factors where data was available for more than 5% of the cohort were included, therefore 17q gain was not further analysed because of small number. *Age was dichotomised according to the methods described by London et al.[15] and the analysis to identify an optimal age cut-off was repeated excluding those 3,666 COG participants (data not given). This analysis confirmed the findings of London et al.[15] and supported an optimal cut-off between 15 and 19 months. The consensus of the INRG task force was a cut-off of 18 months (547 days), though for people with diploid tumours, metastases, and without *MYCN* amplification 12 months (365 days) was used.
Outcomes examined	Primary endpoint: 5-year EFS, defined as time from diagnosis until first occurrence of relapse, progression, secondary malignancy, death, or time of last contact if none of these events occurred. Secondary endpoint: Overall survival (OS) All prognostic factors were analysed as dichotomous variables. Kaplan-Meier curves were examined for each factor and Cox proportional hazards regression models used to identify the most highly significant variable to create a given split or "branch" in the survival tree. The prognostic significance of the remaining factors was then tested within these branched subgroups. This process was repeated until the sample size was too small, or until no further statistically significant variables were found.
Results/outcomes	Median follow-up was 5.2 years. Overall for the full cohort: 5 year EFS 63% +/-1% and OS 70% +/-1% Characteristics for the full cohort are presented in Table i. Within the overall cohort (n=8,800), the presence of metastases was the most significant prognostic factor: • 5 year EFS for INSS non-stage 4 (1, 2, 3 or 4S) was 83% vs. 35% for stage 4 • 5 year OS for INSS non-stage 4 (1, 2, 3 or 4S) was 91% vs. 42% for stage 4 Within favourable non-stage 4 (n=5,131), histological category was the most significant prognostic indicator: • 5 year EFS for GN maturing or GNB intermixed was 97%, vs. 83% for NB and GNB nodular • 5 year OS for GN maturing or GNB intermixed was 98%, vs. 90% for NB and GNB nodular Within NB and GNB nodular (n=4,970), *MYCN* amplification was the most significant prognostic indicator: • 5 year EFS for no *MYCN* amplification was 87%, vs. 46% for *MYCN* amplification • 5 year OS for no *MYCN* amplification was 95%, vs. 53% for *MYCN*

	amplification Following this, INSS stage (1, 2, 3 or 4S) was the most significant prognostic indicator. Subsequent tree regression analyses in groups with and without *MYCN* amplification are not reported here. Within non-favourable stage 4 (n=3,425), age was the most significant prognostic indicator: • 5 year EFS for age <547 days (18 months) was 63% vs. 23% for age ≥547 days • 5 year OS for age <547 days (18 months) was 68% vs. 31% for age ≥547 days For age <547 days *MYCN* amplification was the next most significant prognostic indicator. For age ≥547 days serum ferritin was the next most significant prognostic indicator. Subsequent tree regression analyses in these groups are not reported here. <u>INRG classification system</u> The overall consensus INRG classification schema included 7 factors that were statistically significant and clinically relevant: • INRG stage (as presented in Monclair et al.) • age • histologic classification • grade of tumour differentiation • *MYCN* status • 11q aberrations • tumour cell ploidy This schema produced 16 clinically different pre-treatment risk groups from very low to high risk (presented in Table ii). The proportion of patients in these pre-treatment risk groups with arbitrary 5-year EFS cut-offs applied: • Very low risk group: 28.2% with 5-year EFS >85% • Low risk group: 26.8% with 5-year EFS >75 to ≤85% • Intermediate risk group: 9.0% with 5-year EFS >50 to ≤75% • High risk group: 36.1% with 5-year EFS <50%
Comments	Retrospective data analysis increases the possibility of inaccurate risk estimates. There is the possibility of missing data, and that other factors are confounding analyses (e.g. analyses were not adjusted for other variables, such as treatment

	approach). The INRG classification system aims to ensure that children diagnosed with neuroblastoma are stratified into homogenous, pre-treatment risk groups. The INRG explicitly do not recommend that treatment be assigned according to the four broad risk categories of very low, low, intermediate and high risk, but say that it will be critical to report outcomes for patients assigned to each of the 16 pre-treatment risk groups. Prospective studies assessing prognosis when patients are stratified according to the INRG classification system are needed.

Table i: Clinical and genetic characteristics of the INRG analytic cohort (n=8,800)[16]

Prognostic factor	Patients (n, %)	5-year EFS % (log rank p <0.0001 or specified)	5-year OS % (log rank p <0.0001 or specified)	EFS Hazard ratio (95% CI)
Age <365 days	3,734 (42%)	84%	91%	3.6 (3.3 to 4.0)
Age ≥365 days	5,066 (58%)	49%	55%	
Age <547 days	4,773 (54%)	82%	88%	3.7 (3.4 to 4.0)
Age ≥547 days	4,027 (46%)	42%	49%	
INSS 1, 2, 3, 4S	5,131 (60%)	83%	91%	5.2 (4.8 to 5.7)
INSS stage 4	3,425 (40%)	35%	42%	
Evans I, II, III, IVS	2,022 (63%)	86%	91%	6.6 (5.8 to 7.6)
Evans stage IV	1,177 (37%)	31%	36%	
Ferritin <92 ng/ml	2,170 (50%)	81%	87%	3.6 (3.2 to 4.0)
Ferritin ≥92 ng/ml	2,175 (50%)	46%	52%	
LDH <587 U/L	2,586 (50%)	77%	85%	2.4 (2.2 to 2.7)
LDH ≥587 U/L	2,592 (50%)	53%	58%	
Histologic classification (INPC)				
Favourable	2,724 (64%)	89%	95%	6.6 (5.7 to 7.5)
Unfavourable	1,536 (36%)	40%	49%	
Diagnostic Category (INPC)				
1=NB, stroma poor	3,657 (90%)	64%	71%	
2=GNB, intermixed, stroma rich	144 (3%)	95%	96%	
3=GNB, well differentiated, stroma rich	38 (1%)	80%	79%	
4=GNB nodular				

composite (2&3) vs. (1&4)	232 (6%)	53%	68%	4.7 (2.8 to 7.8)
INPC Grade of NB differentiation Differentiating Undifferentiated	 518 (16%) 2,759 (84%)	 83% 63%	 89% 72%	 2.5 (2.0 to 3.3)
MKI (INPC) Low, intermediate High	 2,690 (87%) 393 (13%)	 74% 37%	 82% 44%	 3.2 (2.8 to 3.8)
MYCN not amplified *MYCN* Amplified	5,947 (84%) 1,155 (16%)	74% 29%	82% 34%	4.1 (3.8 to 4.5)
Ploidy >1 (hyperdiploid) ≤1 (diploid, hypodiploid)	 2,611 (71%) 1,086 (29%)	 76% 55%	 82% 60%	 2.3 (2.0 to 2.6)
11q normal 11q aberration	844 (79%) 220 (21%)	68% 35%	79% 57%	2.3 (1.9 to 2.9)
1p normal 1p aberration	1,659 (77%) 493 (23%)	74% 38%	83% 48%	3.2 (2.8 to 3.8)
No 17q gain 17q gain	187 (52%) 175 (48%)	63% 41% (p=0.0006)	74% 55% (p=0.0009)	1.7 (1.3 to 2.3)
Year of diagnosis ≥1996 <1996	 4,493 (51%) 4,307 (49%)	 69% 59%	 76% 66%	 1.4 (1.2 to 1.4)
Initial treatment Observation, surgery, standard chemo Intensive multimodal	 4,515 (68%) 2,170 (32%)	 79% 34%	 86% 41%	 4.1 (3.8 to 4.4)

Table ii: INRG consensus pre-treatment classification schema[16]

INRG Stage	Age (months)	Histologic category	Grade of differentiation	*MYCN* status	11q aberration	Ploidy	Pretreatment risk group
L1/L2		GN maturing or GNB intermixed					A: Very low
L1		Any, except GN maturing or GNB intermixed		Not amplified			B: Very low
				Amplified			K: High

L2	<18 months	Any, except GN maturing or GNB intermixed		Not amplified	No		D: Low
					Yes		G:Intermediate
	≥18 months	GNB nodular or NB	Differentiating	Not amplified	No		E: Low
					Yes		H:Intermediate
			Poorly or undifferentiated	Not Amplified			H:Intermediate
				Amplified			N: High
M	<18 months			Not Amplified		Hyperdiploid	F: Low
	<12 months			Not Amplified		Diploid	I: Intermediate
	12 to <18			Not Amplified		Diploid	J: Intermediate
	<18 months			Amplified			O: High
	≥18 months						P: High
MS	<18 months			Not amplified	No		C: Very low
					Yes		Q: High
				Amplified			R: High

Appendix number	3
Relevant criteria	2
Publication details	Monclair et al. The International Neuroblastoma Risk Group (INRG) Staging System: An INRG Task Force Report. 2009; 27(2): 298-303.[17]
Study details	Retrospective cohort study Country/Setting: Patient data from the International Society of Pediatric Oncology Europe Neuroblastoma Group (SIOPEN). Since 1994 SIOPEN has classified loco-regional tumours as resectable or unresectable depending on the presence of "surgical risk factors" (features detected on imaging that make safe, complete tumour excision impracticable at the time of diagnosis). Their previous study (LNESG1) evaluated how surgical risk factors detected on imaging was associated with resection outcomes and surgical complications.[18] The SIOPEN principle for stratifying patients with loco-regional tumors by imaging features was adopted by the INRG Task Force at conference in Canada in 2005, and used in the design of the INRG Staging System (INRGSS). One German study (Simon et al.[19] outlined below) has also investigated the prognostic value of these IDRF by retrospectively applying them to a trial population with neuroblastoma. Monclair et al. here report aiming to validate the findings of the German study using data from SIOPEN, which is the only data available in the INRG database able to validate the significance of image-defined risk factors (IDRFs) and INGRSS.

Study objectives	The INRG aimed to develop a consensus approach for pre-treatment risk stratification of neuroblastoma to help facilitate the comparison of risk-based clinical trials conducted in different parts of the world by defining homogenous pre-treatment patient cohorts. The companion publication by Cohn et al. outlines the other prognostic factors in the INRG consensus classification schema. This report outlines the INRGSS according to the presence of image-defined risk factors.
Inclusions/study population	661 patients from SIOPEN with INSS stage 1, 2 and 3 disease with known data for IDRFs.
Exclusions	None reported
Intervention/test and comparator	Not applicable
Prognostic variables examined	Presence or absence of IDRF. The INRGSS outlines 4 stages: • L1: Localized tumour not involving vital structures as defined by the list of image-defined risk factors* and confined to one body compartment • L2: Loco-regional tumour with presence of one or more image-defined risk factors* • M: Distant metastatic disease (except stage MS) • MS: Metastatic disease in children younger than 18 months with metastases confined to skin, liver, and/or bone marrow (<10% of total cells) * The detailed list of the specific consensus-defined IDRF in neuroblastic tumours is not reported here. Alongside IDRFs it includes three conditions to be recorded, but not considered IDRF: multifocal primary tumours, pleural effusion, and ascites.
Outcomes examined	Primary endpoint: 5-year EFS, defined as time from diagnosis until first occurrence of relapse, progression, secondary malignancy, death, or time of last contact if none of these events occurred. Secondary endpoint: OS Univariate analyses were performed to assess the prognostic ability of INRGSS. Kaplan-Meier curves were generated, and curves were compared using log-rank test, with p values <0 .05 considered statistically significant. Outcomes were reported by INRGSS and INSS stage.
Results/outcomes	• 332/661 patients (50%) had no IDRF (i.e. INGRSS L1): 79% of INSS stage 1, 55% of INSS stage 2, and 6% of INSS stage 3 met criteria for INGRSS L1

	• 329/661 patients (50%) had IDRF (i.e. INGRSS L2): 21% of INSS stage 1, 45% of INSS stage 2, and 94% of INSS stage 3 met criteria for INGRSS L2 474/661 had available outcome data • 5-year EFS for INGRSS L1 (n=213) was 90% (+/-3%) vs. 78% (+/-4%) for L2 (n=261) (p=0.0010) • 5-year OS for INGRSS L1 was 96% (+/-2%) vs. 89% (+/-3%) for L2 (p=0.0068) • 5-year EFS for INSS stage 1 (n=209) was 92% (+/-3%) which was also significantly higher than stage 2 (n=103) (78% +/-6%; p=0.0005) and stage 3 (n=162) (75% +/-5%; p<0.0001). There was no significant difference in EFS for INSS stage 2 and 3 (p=0.6611) • 5-year OS for INSS stage 1 was 98% (+/-2%), 95% (+/-3%) for stage 2, and 84% (+/-4%) (p not reported)
Comments	Retrospective data analysis increases the possibility of inaccurate risk estimates. Not all people in SIOPEN had data available on IDRFs or outcomes. Other factors may be confounding analyses (e.g. treatment approach). The INRG classification system aims to ensure that children diagnosed with neuroblastoma are stratified into homogenous, pre-treatment risk groups. The primary function of the INGRSS is as a component of the INRG pre-treatment risk classification system, and it is not intended to be a substitute for the INSS. Prospective studies assessing the prognostic significance of IDRF, and when patients are stratified according to the broader INRG classification system, are needed.

Appendix number	4
Relevant criteria	2
Publication details	Simon et al. Review of Image Defined Risk Factors in Localized Neuroblastoma Patients: Results of the GPOH NB97 Trial. 2008; 50: 965-969.[19]
Study details	Retrospective cohort study (data collected prospectively as part of a trial but retrospective analysis of outcomes according to radiology reports) Country: Germany Setting: Patients in the NB97 trial, diagnosed Oct 1996 to Dec 2003.
Study objectives	Retrospective review of the impact of IDRF (as adopted by the INRG Task Force at conference in Canada in 2005) on extent of tumour resection, surgery-associated complications and prognostic value.
Inclusions/study population	366 patients with localised neuroblastoma unequivocal information available on the presence or absence of IDRF at the time of diagnosis.

Exclusions	Patients with *MYCN* amplification
Intervention/test and comparator	Not applicable
Prognostic variables examined	Presence or absence of IDRF as defined by the INRG Task Force at conference in Canada in 2005 (outlined in Monclair et al.[17]) Multivariate Cox regression analysed the prognostic significance of IDRF, in addition to INSS stage, age, tumour crossing the midline, and treatment protocol (observation or standard risk).
Outcomes examined	• 3 year EFS (time from diagnosis to relapse, progression or death, or date of last examination) • 3 year OS • Extent of tumour resection • Surgery-associated complications (not further reported here)
Results/outcomes	• 227/366 patients (62%) had no IDRF (i.e. INGRSS L1): 59% of whom were INSS stage 1, 28% were INSS stage 2, and 13% were INSS stage 3 • 139/366 patients (38%) had any IDRF (i.e. INGRSS L2): 19% of whom were INSS stage 1, 35% were INSS stage 2, and 46% were INSS stage 3 The presence of IDRFs was a poor prognostic indicator for EFS but not OS: • 3 year EFS: 86% (+/-2%) without IDRF vs. 75% (+/-2%) with any IDRF (p=0.010) • 3 year OS was no different: 98% (+/-1%) without IDRF vs. 96% (+/-2%) with any IDRF (p=0.462) According to INSS: • 3 year EFS: 93% (+/-2%) for INSS 1 vs. 78% (+/-4%) for INSS 2 vs. 69% (+/-5%) for INSS 3 (p<0.001) • 3 year OS: 98% (+/-1%) for INSS 1 vs. 99% (+/-1%) for INSS 2 vs. 94% (+/-3%) for INSS 3 (p=0.056) By subgroup • In a subgroup of 260 who did not receive chemotherapy, 3 year EFS was 90% (+/-2%) for those without IDRF vs. 72% (+/-5%) with any IDRF (p=0.001). OS in this group was no different: 99% (+/-1%) without and 98% (+/-2%) with IDRF (p=0.572) • In a subgroup of 98 aged >18 months IDRF had no effect on 3 year EFS (80% +/-5% without and 69% +/-8% with IDRF; p=0.304) or OS (96% +/-3% without and 91% +/-5% with IDRF; p=0.312) Other treatment outcomes • Complete resection during the primary operation was achieved in 156/227 without IDRF (69%) vs. 43/149 (31%) with IDRF (p<0.001). Combining the results of first and subsequent operations (after chemotherapy), complete resection was achieved in 81% without IDRF and 51% with IDRF (p<0.001) • In the subgroup of 260 who did not receive chemotherapy, complete

	resection was achieved in 146/189 without IDRF (77%) vs. 36/71 (51%) with IDRF (p<0.001) after the first operation, and 15/189 (8%) vs. 6/71 (8%) at second operation. In multivariate analysis presence of IDRF was not an independent risk factor for an event (p=0.557), nor were age or treatment (standard risk vs. observation). INSS stage 2 or 3 were associated with significantly increased risk of an event compared with stage 1 (p<0.001).
Comments	Retrospective analysis of a trial population, where other factors may be confounding analyses. Subgroup analyses, particularly by age, are also small. Prospective studies evaluating the significance of IDRFs are needed.

Appendix number	5
Relevant criteria	13
Publication details	Hisashige A for the NBS Evaluation Group. Effectiveness of nationwide screening program for neuroblastoma in Japan. Global Journal of Health Science. 2014; 6(4): 94-106.[12]
Study details	Retrospective cohort study (non-randomised controlled study) Country: Japan. Setting: 25 study areas (prefectures) from eight districts where neuroblastoma screening was performed using high performance liquid chromatography (HPLC) between 1984 and 1997.
Study objectives	To compare the incidence of and mortality from neuroblastoma after six months of age between children screened using HPLC, and non-screened children from the same population (i.e. children from the same areas who did not participate in the screening programme).
Inclusions	Children born after the change of screening test from qualitative methods to HPLC in each of 25 study areas (from eight districts), from its earliest introduction in January 1984 to December 31 1997.
Exclusions	Children born in three study areas where a double screening program was started during the study period (Sapporo City, Hokkaido 6 and 14 months; Miyagi Prefecture 6 and 18 months; and Kyoto Prefecture 6 and 18 months). 25 of 47 study areas were selected based on the availability of information on cases (as outlined under the population). This represented approximately half of the children born in Japan during the study period.
Population	Screened (n=3,705,670) and non-screened children (n=603,900) as identified from the participant list of the screening program in each of the 25 areas (total

	4,309,570 births during the study period). Cases of neuroblastoma were identified by contacting hospitals in each area and confirming with patient data sources including medical records of the National Medical Aid Program for Specific Chronic Pediatric Diseases, the Japan Children's Cancer Registry, the prefecture Cancer Registry, and the prefecture Registry of Childhood Malignancies. For each identified case investigators filled out the standardised patient report form, including data on: • gender • birthplace • information source • screened or non-screened at 6 months • screening results • date of diagnosis • method of detection (screening or clinical symptoms) • clinical stage at diagnosis (INSS) • status alive or dead on December 31 1998, and date of death • cause of death (defined as death from neuroblastoma; related to neuroblastoma treatment; from other cancers; or death from other causes) Incident cases were defined according to the Evans staging system (Evans et al. 1971) or INSS. Neuroblastoma mortality was defined as death from neuroblastoma or related treatment. There were a total 22,875,800 person-years of follow-up (19,532,468 for the screened group, 3,343,332 for the non-screened group). Mean observation time was from 6 months to 5.3 years of age.
Intervention/test	Neuroblastoma screening at 6 months by quantitative measurement of vanillylmandelic acid (VMA) and homovanillic acid (HVA) using high performance liquid chromatography (HPLC). This method was introduced nationally from 1990 onwards to replace the less accurate qualitative methods of VMA detection using spot test or thin layer chromatography (TLC).
Comparator	No screening (children from the same areas not participating in screening).
Results/outcomes	<u>Incidence</u> Total 767 incident cases after 6 months during the study period: 708 in the screened and 59 in the non-screened group. The incidence according to age group, with incidence rate (IR) per million person years, and Incidence Rate Ratio (IRR):

	• Incidence from 6 months to 1 year: 602 screened vs. 20 non-screened; IR 325.92 screened vs. 66.80 non-screened; IRR significantly higher in screened children: 4.88, 95% CI 3.13 to 7.62 • Incidence from 1 to 4 years: 98 screened vs. 36 non-screened; IR: 8.14 screened vs. 17.93 non-screened; IRR significantly less in screened children: 0.45, 95% CI 0.31 to 0.67 • Incidence >5 years: 8 screened vs. 3 non-screened; IR: 1.55 screened vs. 3.74 non-screened; IRR no different: 0.48, 95% CI 0.11 to 1.61 By stage at diagnosis: • Incidence of early stage (1, 2, 4S) from 6 months to 1 year: 472 screened vs. 8 non-screened; IR 255.54 vs. 26.72; IRR significantly higher in screened children: 9.56, 95% CI 4.76 to 19.23 • Incidence of advanced stage (3,4) from 6 months to 1 year: 130 screened vs. 12 non-screened; IR 70.38 vs. 40.08; IRR no different: 1.76, 95% CI 0.97 to 3.17 • Incidence of early stage from 1 to 4 years: 31 screened vs. 8 non-screened; IR 2.58 vs. 3.98; IRR no different: 0.65, 95% CI 0.30 to 1.41 • Incidence of advanced stage from 1 to 4 years: 67 screened vs. 28 non-screened; IR 5.57 vs. 13.94; IRR significantly less in screened children: 0.40, 95% CI 0.26 to 0.62 • Incidence of early stage >5 years: 0 in both screened and non-screened groups • Incidence of advanced stage >5 years: 8 screened vs. 3 non-screened; IR 1.55 vs. 3.74; IRR no different 0.42, 95% CI 0.11 to 1.61 <u>Mortality</u> A total 87 children with neuroblastoma died by the end of 1998. After exclusion of 14 cases detected clinically before 6 months of age, 7 who died from other causes, and 1 with no data available, the total mortality from neuroblastoma after 6 months of age was 66: 49 in the screened vs. 17 in the non-screened group The mortality rate (MR) above 6 months per million person years was 2.46 in the screened vs. 4.50 in the non-screened group; the mortality rate ratio (MRR) was significantly less in screened children: 0.547, 95% CI 0.306 to 0.976. Age specific mortality rates: • Age at death 6 months to one year: 2 screened and 1 non-screened; MR

	1.08 vs. 3.31; MRR no different: 0.326, 95% CI 0.030 to 3.595 • Age at death 1 to 3 years: 31 screened and 12 non-screened; MR 3.11 vs. 7.51; MRR significantly less in screened children: 0.415, 95% CI 0.212 to 0.810 • Age at death >4 years: 16 screened and 4 non-screened; MR 2.17 vs. 2.46; MRR no different: 0.880, 95% CI 0.255 to 3.040
Comments	Retrospective cohort study. The non-randomised study design increases the possibility of selection bias. The study has compared screened and non-screened children from the same area, rather than previous cohorts that have compared screened regional areas with non-screened control areas. This may remove some possible bias (e.g. differences in environmental and population characteristics or healthcare services, including diagnostic and treatment approaches); however, there may be differences between those who have and have not chosen to participate in screening which may be confounding the results. Characteristics of screened and non-screened participants are not reported in the study. There is also some possibility for recall/information bias. Though data collection was comprehensive it came from variable sources. For example, information on neuroblastoma cases came from hospitals in 72% of cases and registration databases for the remainder, and screening participation for these cases came from the participant list for 74%, parental interview in 24% and medical records in 2%. The study specifically examined the HPLC test, and findings may not be comparable to other studies that have evaluated other methods such as TLC, or a secondary test (e.g. gas chromatography). This study also specifically evaluated screening at 6 months only, compared to no screening, and cannot inform on the value of testing at other ages.

Appendix number	6
Relevant criteria	13
Publication details	Hiyama E et al. Effectiveness of screening for neuroblastoma at 6 months of age: a retrospective population-based cohort study. Lancet. 2008; 371:1173-80.[11]
Study details	Retrospective cohort study (before-after study) Country: Japan. Setting: nationwide study comparing children born in three periods: 1980 to 1983

	(before screening), 1986 to 1989 (qualitative screening), and 1990 to 1998 (quantitative screening using HPLC).
Study objectives	To assess the effectiveness of screening for neuroblastoma at 6 months, by investigating the incidence and mortality in children younger than 6 years of age for children born in the three periods.
Inclusions	Children born nationwide between 1980 and 1998
Exclusions	Children born in 1984 and 1985, to exclude the crossover period when nationwide screening was being introduced.
Population	Total study population 22,289,695 comprising children born in 1980 to 1983 before nationwide screening (n=6,130,423); children born in 1986 to 1989 during nationwide qualitative screening (n=5,290,412); and children born in 1990 to 1998 during nationwide quantitative screening (n=10,868,860). 77.4% of the 1986 to 1989 birth cohort (n=4,092,759) received qualitative screening. 86.0% of the 1990 to 1998 birth cohort (n=9,342,132) received quantitative screening. Neuroblastoma incidence was identified from cancer registries of the Japanese Society of Paediatric Surgeons and the Japanese Society of Paediatric Oncology. These databases were cross-referenced against the Japanese Infantile Neuroblastoma Cooperative Study Group. The researchers also requested major hospitals to supply clinical data for cases identified during the pre-screening and qualitative cohorts. Mortality and cause of death was identified from the database and verified by death certificates provided by the Ministry of Health, Labour and Welfare. Case matching showed that 62.5% of neuroblastoma deaths were registered in the database (527/843). This included 206/329 deaths for the 1980 to 1983 cohort, 129/201 for 1986 to 1989, and 192/313 for 1990 to 1998.
Intervention/test	Nationwide qualitative screening using TLC or quantitative screening using HPLC
Comparator	Before period of no nationwide screening
Results/outcomes	<u>Incidence</u> Total 3181 cases of neuroblastoma diagnosed before the age of 6 years for all children in the cohort: 443 in the pre-screening cohort, 713 in the qualitative cohort, and 2025 in the quantitative cohort. 60% of tumours were detected by qualitative screening (430/713; 124 false negatives) and 76% were detected by quantitative screening (1537/2025; 244

	false negatives) during the respective periods. 266 of all cases were diagnosed prior to 6 months: 60 in the pre-screening, 71 in the qualitative screening, and 135 in the quantitative screening cohort. Total cumulative incidence per 100,000 births (calculated using the registration rate of 62.5%) for the three periods, and relative risk (RR with 95% CI) compared with pre-screening: • Pre-screening: 443 cases; cumulative incidence 11.56 per 100,000 • Qualitative screening: 713 cases, cumulative incidence 21.56, RR 1.87 (1.66 to 2.10) • Quantitative screening: 2025 cases, cumulative incidence 29.80, RR 2.58 (2.33 to 2.86) Specifically for cases diagnosed between 6 months to 6 years, compared with pre-screening (excluding diagnoses prior to 6 months): • Pre-screening: 383 cases; cumulative incidence 10.00 per 100,000 • Qualitative screening: 642 cases, cumulative incidence 19.42, RR 1.94 (1.71 to 2.21); among screened cumulative incidence 21.66, RR 2.17 (1.90 to 2.47); among non-screened cumulative incidence 11.76, RR 1.18 (0.92 to 1.47) • Quantitative screening: 1890 cases, cumulative incidence 27.82, RR 2.78 (2.50 to 3.11); among screened cumulative incidence 30.50, RR 3.05 (2.74 to 3.41); among non-screened cumulative incidence 11.42, RR 1.14 (0.91 to 1.40) By stage at diagnosis, compared with pre-screening: • Pre-screening: stage 1-3 CI 3.99, stage 4 cumulative incidence 5.21, stage 4S cumulative incidence 0.21 • Qualitative screening: stage 1-3 cumulative incidence 13.28, RR 3.32 (2.78 to 4.01); stage 4 cumulative incidence 4.69, RR 0.90 (0.73 to 1.11); stage 4S cumulative incidence 0.78, RR 3.77 (1.89 to 9.36) • Quantitative screening: stage 1-3 cumulative incidence 21.76, RR 5.44 (4.63 to 6.47); stage 4 cumulative incidence 4.27, RR 0.82 (0.69 to 0.98); stage 4S cumulative incidence 1.49, RR 7.11 (3.79 to 16.9) When subanalysed by uptake of screening in the two periods, for qualitative screening the increase in incidence of stages 1 to 3 was significant for both screened and non-screened children; for stage 4 there was no significant difference for either screened or non-screened; and for 4S the significant increase

	was only among those who were screened. For quantitative screening, the significant difference in incidence of all stages (increased incidence for stages 1 to 3 and 4S; decreased incidence for stage 4) was only seen among the screened children. For non-screened there was no significant difference from pre-screening. <u>Mortality</u> Total cumulative mortality rate (MR) per 100,000 births (adjusted for registration rate in the database) and RR (95% CI) compared with pre-screening: • Pre-screening: 206 deaths, mortality rate 5.38 • Qualitative screening: 129 deaths, MR 3.90, RR 0.73 (0.58 to 0.90) • Quantitative screening: 192 deaths, MR 2.83, RR 0.53 (0.42 to 0.63) Specific mortality in those between ages 6 months and 6 years (excluding deaths prior to 6 months): • Pre-screening: 192 deaths, mortality rate 5.01 • Qualitative screening: 118 deaths, MR 3.57, RR 0.71 (0.56 to 0.89) • Quantitative screening: 174 deaths, MR 2.56, RR 0.51 (0.42 to 0.63) By screening participation: • Qualitative screening: screened: 82 deaths, MR 3.21, RR 0.64 (0.49 to 0.82); non-screened: 36 deaths, MR 4.81, RR 0.96 (0.6 to 1.34) • Quantitative screening: screened: 128 deaths, MR 2.19, RR 0.44 (0.35 to 0.54); non-screened: 46 deaths, MR 4.79, RR 0.96 (0.68 to 1.30)
Comments	Retrospective cohort study using a before-after design. There was no specific control group of non-screened children. Subanalyses were conducted comparing incidence and mortality between those who did and did not participate in screening, though selection bias may be confounding these analyses. Changes in incidence and mortality over time could reflect changes in diagnostic methods, treatments and healthcare resources, rather than the sole effect of screening. This study specifically evaluated screening at 6 months only, comparing qualitative and quantitative methods with the period before screening. It cannot inform the value different screening methods (e.g. single vs. double stage), or of testing at ages other than 6 months.

References

1. Neuroblastoma [Internet]. Children with cancer UK; [cited 2015 Jan 29]. Available from: http://www.childrenwithcancer.org.uk/neuroblastoma.
2. RCR. Recommendations for cross-sectional imaging in cancer management, second edition: paediatric neoplasms. London: The Royal College of Radiologists, 2014. Available from: https://www.rcr.ac.uk/docs/radiology/pdf/BFCR%2814%292_22_Paediatric.pdf.
3. CRUK. Neuroblastoma [Internet]. Cancer Research UK; [updated 2014 Sep 5; cited 2015 Jan 29]. Available from: http://www.cancerresearchuk.org/about-cancer/cancers-in-general/cancer-questions/neuroblastoma.
4. NCI. Neuroblastoma screening (PDQ): summary of evidence [Internet]. National Cancer Institute; [updated 2014 Feb 6; cited 2015 Jan 29]. Available from: http://www.cancer.gov/cancertopics/pdq/screening/neuroblastoma/HealthProfessional/.
5. Heck JE, Ritz B, Hung RJ, et al. The epidemiology of neuroblastoma: a review. Paediatr Perinat Epidemiol. 2009;23(2):125-43.
6. Brisse HJ, McCarville MB, Granata C, et al. Guidelines for imaging and staging of neuroblastic tumors: consensus report from the International Neuroblastoma Risk Group Project. Radiology. 2011;261(1):243-57.
7. Beiske K, Burchill SA, Cheung IY, et al. Consensus criteria for sensitive detection of minimal neuroblastoma cells in bone marrow, blood and stem cell preparations by immunocytology and QRT-PCR: recommendations by the International Neuroblastoma Risk Group Task Force. Br J Cancer. 2009;100(10):1627-37.
8. Yalcin B, Kremer LC, Caron HN, et al. High-dose chemotherapy and autologous haematopoietic stem cell rescue for children with high-risk neuroblastoma. Cochrane Database Syst Rev. 2013;8:CD006301.
9. Woods WG, Gao R-U, Shuster JJ, et al. Screening of infants and mortality due to neuroblastoma. NEJM. 2002;346:5.
10. Schilling FH, Spix C, Berthold F, et al. Neuroblastoma screening at one year of age. NEJM. 2002;346:6.
11. Hiyama E, Iehara T, Sugimoto T, et al. Effectiveness of screening for neuroblastoma at 6 months of age: a retrospective population-based cohort study. Lancet. 2008;371(9619):1173-80.
12. Hisashige A, Group ftNE. Effectiveness of nationwide screening program for neuroblastoma in Japan. Global Journal of Health Science. 2014;6:13.
13. Riley RD, Burchill SA, Abrams KR, et al. A systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma. Health Technol Assess. 2003;7(5).
14. Haber M, Smith J, Bordow SB, et al. Association of high-level MRP1 expression with poor clinical outcome in a large prospective study of primary neuroblastoma. J Clin Oncol. 2006;24(10):1546-53.
15. London WB, Castleberry RP, Matthay KK, et al. Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the Children's Oncology Group. J Clin Oncol. 2005;23(27):6459-65.

16. Cohn SL, Pearson AD, London WB, et al. The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. J Clin Oncol. 2009;27(2):289-97.
17. Monclair T, Brodeur GM, Ambros PF, et al. The International Neuroblastoma Risk Group (INRG) staging system: an INRG Task Force report. J Clin Oncol. 2009;27(2):298-303.
18. Cecchetto G, Mosseri V, De Bernardi B, et al. Surgical risk factors in primary surgery for localized neuroblastoma: the LNESGI study of the European International Society of Pediatric Oncology neuroblastoma group. J Clin Oncol. 2005;23(23):7.
19. Simon T, Hero B, Benz-Bohm G, et al. Review of image defined risk factors in localized neuroblastoma patients: Results of the GPOH NB97 trial. Pediatr Blood Cancer. 2008;50(5):965-9.
20. Schleiermacher G, Mosseri V, London WB, et al. Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. Br J Cancer. 2012;107(8):1418-22.
21. Moroz V, Machin D, Faldum A, et al. Changes over three decades in outcome and the prognostic influence of age-at-diagnosis in young patients with neuroblastoma: a report from the International Neuroblastoma Risk Group Project. Eur J Cancer. 2011;47(4):561-71.