NHS England

Evidence review: The safety of proton beam therapy for childhood tumours

NHS England

Evidence review: The safety of proton beam therapy for childhood tumours

First published: June 2018

Updated: Not applicable

Prepared by: Solutions for Public Health (SPH) on behalf of NHS England Specialised Commissioning

Contents

1 Introduction

Introduction

- While malignancies in childhood and early adult life are not common, they are still one of the leading causes of death in those age-groups.
- The objective of this evidence review is to investigate the safety of proton beam therapy (PBT) in the treatment of malignancy in childhood and early adult life, relative to photon radiotherapy.

Existing guidance from the National Institute for Health and Care Excellence (NICE)

- We found no guidance from NICE about the use of PBT. NICE's interventional procedures programme has not produced guidance on PBT for malignant brain tumours because there is no CE marked device (NICE).

The indication and epidemiology

- This rapid evidence review is concerned with the following malignancies:
  - medulloblastoma
  - ependymoma
  - craniopharyngioma
  - other brain tumours
  - salivary gland tumours
  - retinoblastoma

- Medulloblastoma is the commonest malignant brain tumour in children, with an annual incidence of about two per million. It presents most often in children between the ages of three and eight years, with an abrupt onset of headaches, especially in the morning, nausea and/or vomiting, tiredness and visual movement abnormalities (Hirano et al 2014).

- Ependymoma is a tumour of a type of glial cell called ependymal cells. The commonest site in children is the lining of the fourth ventricle. It accounts for 6% to 10% of intracranial tumours in childhood (Gunther 2015).

- Craniopharyngiomas arise from embryonic tissue associated with the pituitary gland, and occur in the suprasellar region. The incidence is about 1.3 per million person-years (Zacharia et al 2012).

- Malignant salivary gland tumours are rare in children, with an estimated annual incidence of 0.4 per million (Grant 2015).

- Retinoblastoma is a tumour that arises from the immature cells of a retina. Its incidence is 11.8 cases per million children less than 5 years of age, with most cases occurring in children less than two years old (Sethi 2014).

Standard treatment and pathway of care

- Standard pathways of care vary between the tumours covered by this rapid evidence review. In most cases, initial treatment is with surgery, followed in some cases by

chemotherapy and/or radiotherapy to improve the prospects of cure. Retinoblastomas are however usually treated with laser therapy, cryotherapy or brachytherapy.

The intervention

- Radiotherapy uses radiation to destroy malignant tissue while minimising damage to adjacent normal tissue. PBT uses a high-energy beam of protons as treatment, rather than the high-energy X-ray photons used in standard radiotherapy for patients with cancer.

- The only NHS proton beam therapy centre is at the Clatterbridge Cancer Centre NHS Foundation Trust. It delivers a low-energy proton therapy specifically for patients with eye tumours. Patients who require proton therapy for other tumours may be referred overseas via the NHS Proton Overseas Programme.

- Two further facilities are being built in Manchester and London, and are expected to open in 2018 and 2020 respectively.

Rationale for use

- PBT is intended to deliver highly targeted radiation to the tumour with less collateral damage.

2 Summary of results

- Eleven papers matching the PICO were included in this review.

Medulloblastoma

- Paulino et al 2018 reported results in 84 children with medulloblastoma. They reported rates of hearing loss of grade 3 or 4 on the SIOP Boston scale[1]; after PBT, these were 15/75 (20%), and after photon radiotherapy (PRT) 21/91 (23%), p=0.63. The authors report three other measures of hearing loss, but none showed a significant difference in its incidence between the participants treated with PBT and PRT.

- Eaton et al 2015 reported 77 children with medulloblastoma treated with craniospinal radiation. Adjusted odds ratios were 0.13 for hypothyroidism (95% CI 0.04 to 0.41, p<0.001), 0.06 for sex hormone deficiency (95% CI 0.01 to 0.55, p=0.013) and 0.30 for endocrine replacement therapy (95% CI 0.09 to 0.99, p=0.047). For participants' height, the standard deviation score parameter estimate was 0.89 (indicating greater height with PBT, 95% CI 0.24 to 1.54, p=0.008).

- Eaton et al 2015's results are not reliable, because of biases in age, diagnostic testing and acceptance of treatment between the two groups. Differences in the timing and purpose of data collection may also have introduced bias.
- Hirano et al 2014 published a health economic model of PBT versus PRT for

[1] A hearing loss scale: grade 0= ≤20dB loss at all frequencies, grade 1 = > 20dB sensorineural hearing loss (SNHL) above 4 kHz, Grade 2 = >20dB SNHL at 4 kHz, Grade 3 = >20 dB SNHL at or above 2 kHz, Grade 4 = Grade 2 = >40 dB SNHL at or above 2 kHz.

medulloblastoma, considering only the risk of hearing loss and its impact on quality of life. Three different measures of quality of life were used: EQ-5D[2]: (£16,100/quality-adjusted life-year (QALY)), HUI3[3] (£8710/QALY) and SF-6D[4] (£14,900/QALY).

- These costs per QALY are well below the threshold of acceptable value of money for the NHS, appearing to indicate that the extra costs of PBT are justified. However, hearing loss rates supported by modern evidence lie outside the sensitivity ranges used by Hirano et al 2014, casting doubt on the reliability of their conclusions.

Ependymoma

- Sato et al 2017 reported a study involving 79 children with intracranial ependymoma. Toxicity rates after PBT were 3/41 (7.3%), after PRT they were 5/38 (13.2%), (χ^2 = 0.237, p=0.626).

- Gunther et al 2015 reported MRI abnormalities with associated symptoms in 72 children with ependymoma. In those receiving PBT, 4/37 (11%) had abnormalities with symptoms, compared with 3/35 (8.6%) after PRT (χ^2 = 0.006, p=0.938).

Craniopharyngioma

- Bishop et al 2014's study included 52 children with craniopharyngioma. The authors reported several adverse effects of treatment, though none showed a significant difference in rates between participants receiving the two treatments:

    - Vascular morbidity, including moyamoya, stroke, and vessel malformations: PBT 2/21 (10%), PRT 3/31 (10%), p=1.0
    - Visual morbidity: PBT 1/21 (5%), PRT 4/31 (13%), p=0.637
    - Hypothalamic obesity: PBT 4/21 (19%), PRT 9/31 (29%), p=0.523
    - Endocrinopathy; PBT 16/21 (76%), PRT 24/31 (77%), p=1.0

Salivary gland tumours

- Grant al 2015 published a small study of 24 children with malignant salivary gland tumours. They report rates of several local adverse effects:

    - Dermatitis: PBT 7/13 (54%), PRT 6/11 (55%), p=1.0.
    - Dysphagia: PBT 0/13 (0%), PRT 3/11 (27%), p=0.08
    - Otitis externa: PBT 1/13 (8%), PRT 2/11 (18%), p=0.58
    - Mucositis: PBT 6/13 (46%), PRT 10/11 (91%), p<0.05

- The reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. Also, correction for multiple testing meant that none of the reported differences is statistically significant.

[2] A standardised instrument for measuring health status
[3] The Health Utilities Index 3, a rating scale used to measure general health status and health-related quality of life
[4] Short form 6 dimension is a measure of health utility

Retinoblastoma

- Sethi et al 2014 reported results in 86 children with retinoblastoma. The rate of second malignancies in the field irradiated by PBT were 0/55 (0%), 95% confidence interval (CI) not reported; after PRT the rate was 4/31 (14%), 95% CI 3% to 31% (p=0.015). Corresponding rates for second malignancies anywhere were [figure not reported]/55 (5%), 95% CI 0% to 21%, and 4/31 (13%), 95% CI 3% to 31% respectively (p=0.120).

- The median length of follow-up for participants treated with PRT was nearly twice that in those who received PBT, but the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group.

Tumours at several sites

- Kahalley et al 2017 reported a study of 123 children with brain tumours. Those who received PBT had no statistically significant decline in intelligence quotient (IQ) (p= 0.130). The children who received PRT has a loss of 1.1 IQ points per year (p= 0.004). However, a comparison of the change in IQ over time between the two groups revealed no significant difference in rates of decline (p= 0.509).

- The authors conclude that "this study does not provide clear evidence that [PBT] results in clinically meaningful sparing of global IQ significantly exceeding that of modern [PRT] protocols."

- Yock et al 2014 analysed the health-related QoL in 120 children with brain tumours. Using the PedsQL[5] Core Module, they reported:
  - mean PedsQL total core score: PBT 75.9, PRT 65.4, unadjusted p=0.002
  - physical summary score PBT 78.4, PRT 68.1, unadjusted p=0.015
  - psychosocial summary score: PBT 74.5, PRT 64.0, unadjusted p=0.001

- This study is affected by biases from family income, socio-economic status, ethnicity and changes in treatment techniques. It is also incorrectly analysed.

- Song et al 2014 reported a study of 43 children with malignancies at various sites, with measures of rates of these adverse effects:
  - Leukopaenia: grade 3[6]: PBT 14/30 (57%), PRT 6/13 (46%); grade 4[7]: PBT 2/30 (7%), PRT 4/13 (31%); p=0.069
  - Anaemia: grade 3[8]: PBT 0/30 (0%), PRT 2/13 (15%), p=0.493
  - Thrombocytopenia: grade 3[9]: PBT 6/30 (20%), PRT 4/13 (31%); grade 4[10]: PBT 1/30 (3%), PRT 3/13 (23%); p=0.012
  - Platelet transfusion: PBT 5/30 (17%), PRT 6/13 (46%), p=0.042
  - Dysphagia: PBT 14/30 (47%), PRT 2/13 (15%), p=0.086
  - Diarrhoea: PBT 0/30 (0%), PRT 3/13 (23%), p=0.023

[5] The PedsQL is a validated assessment of health-related QoL for children with or without chronic health conditions. Scores are from 0 to 100, with 100 representing the best quality of life. PedsQL total scores are in two major sub-domains, physical and psychosocial.
[6] Grade 3: <2000 – 1000/mm^3 (<2.0 – 1.0 x 10^9 /L)
[7] Grade 4: <1000/mm^3 (<1.0 x 10^9 /L)
[8] Hb 6.5 to 8 g/dl.
[9] <1.0 – 0.5 x 109 /L
[10] < 0.5 x 10^9/L (< 500/mm^3)

- Correction for multiple testing of the authors' significance threshold renders all the reported differences non-significant.
- There is a substantial amount of evidence comparing adverse results of PBT and PRT. However, the studies that we found were inconclusive, biased and/or incorrectly analysed. None provided reason to believe that PBT is associated with a lower risk of adverse treatment effects than PRT.
- Randomised trials are needed with appropriate analysis to resolve the uncertainties still present despite the studies included in this review.
- The lack of evidence precludes conclusions about the relative safety of PBT and PRT, about the quantification of safety advantages, about effects on second malignancies or about cost implications of different treatments.

3 Methodology

- The methodology to undertake this review is specified by NHS England in their 'Guidance on conducting evidence reviews for Specialised Commissioning Products' (2016).
- An initial description of the relevant Population, Intervention, Comparison and Outcomes (PICO) to be included in this review was prepared by NHS England's Policy Working Group for the topic (see section 9 for PICO).
- The PICO was used to search for relevant publications in Medline, Embase and Cochrane Library (see section 10 for search strategy).
- The search dates for publications were between 1 January 2008 and 13 April 2018.
- The titles and abstracts of the results from the literature searches were assessed using the criteria from the PICO. Full text versions of papers which appeared potentially useful were obtained and reviewed to determine whether they were appropriate for inclusion. The abstract selection and scoping issues arising from the full paper selection were shared with NHS England in advance of formal approval to proceed with the evidence review.
- Evidence from all papers included was extracted and recorded in evidence summary tables, critically appraised and their quality assessed using National Service Framework for Long Term Conditions (NSF-LTC) evidence assessment framework (see section 7 below).
- The body of evidence for individual outcomes identified in the papers was graded and recorded in grade of evidence tables (see section 8 below).

4 Results

We found ten unrandomised controlled studies comparing the toxicity of PBT and PRT. Of these, two reported results in medulloblastoma (Paulino et al 2018 and Eaton et al 2015), two in ependymoma (Sato et al 2017 and Gunther et al 2015), one each in craniopharyngioma (Bishop et al 2014), retinoblastoma (Sethi et al 2014) and salivary gland tumours (Grant et al

2015), and three in tumours at several sites (Kahalley et al 2017, Yock et al 2014 and Song et al 2014). There was also a health economic study in medulloblastoma (Hirano et al 2014).

Medulloblastoma

Hearing loss

Paulino et al 2018 reported results in 84 children with medulloblastoma treated with craniospinal radiation (PBT 38, PRT 46) and cisplatin-based chemotherapy between 1997 and 2013. They reported rates of hearing loss of grade 3 or 4 on the SIOP Boston scale[11]; after PBT, these were 15/75 (20%), and after PRT 21/91 (23%), p=0.63. The authors report three other measures of hearing loss, but none showed a significant difference in its incidence between the participants treated with PBT and PRT.

Cisplatin is ototoxic (Paken et al 2016). Participants receiving PRT had both higher doses of an ototoxic drug (mean cisplatin dose PBT 281 mg/m^2, PRT 356 mg/m^2, p=0.004) and, according to the authors' modelling, higher radiation doses to their cochleas (mean cochlear radiation dose: PBT 3150 Gy, PRT 3726 Gy, p<0.0001), but there is no reported difference in the risk of hearing loss.

This study therefore indicates that PRT is no more likely to cause hearing loss than PBT.

Endocrinopathy

Eaton et al 2015 reported 77 children with medulloblastoma treated with craniospinal radiation. Forty had PBT, treated in Boston as part of a trial with prospective data collection, and 37 had PRT in Atlanta outside a trial and with retrospective data collection. Multivariable analysis, adjusted for gender, date of diagnosis, histology, location of radiotherapy boost, age at diagnosis and craniospinal radiation dose, reported odds ratios for hypothyroidism of 0.13 (95% CI 0.04 to 0.41, p<0.001), for sex hormone deficiency of 0.06 (95% CI 0.01 to 0.55, p=0.013) and for endocrine replacement therapy of 0.30 (95% CI 0.09 to 0.99, p=0.047). For participants' height, the standard deviation score parameter estimate was 0.89 (indicating greater height with PBT, 95% CI 0.24 to 1.54, p=0.008).

Although Eaton et al 2015 reported several results suggesting lower toxicity after PBT, the results are not reliable:

- The PBT participants were on average more than two years younger than those who received PRT. This may have affected the susceptibility of adjacent tissue to irradiation and biased the study.

- The authors suggest that the differences that they report may be due to biases in diagnostic testing and acceptance of treatment at the two hospitals. For example, the Atlanta participants, treated with PRT, were only tested for growth hormone deficiency when it was clinically suspected, and "*testing may not have been undertaken if the patient/family actively declined the treatment prior to testing. Family willingness to undergo [growth hormone] replacement may have been impacted by social factors such as cost or a fear of the potential impact on tumor recurrence or second malignancy risk. This may have artificially lowered the [growth hormone deficiency] reported, as patients may have had clinical evidence of [growth hormone deficiency] but may not have undergone the confirmatory testing required to make the diagnosis."* By contrast, all

[11] A hearing loss scale: grade 0= ≤20dB loss at all frequencies, grade 1 = > 20dB sensorineural hearing loss (SNHL) above 4 kHz, Grade 2 = > 20dB SNHL at 4 kHz, Grade 3 = > 20 dB SNHL at or above 2 kHz, Grade 4 = Grade 2 = >40 dB SNHL at or above 2 kHz.

Boston participants were recommended to have growth hormone stimulation testing (a confirmatory test) when there was clinical suspicion of the diagnosis.

- Differences in the timing and purpose of data collection may also have introduced bias.

For these reasons, limited reliance can be placed on these results.

Cost utility
Hirano et al 2014 published a health economic model of PBT versus PRT for medulloblastoma in children of 6 years, considering only the risk of hearing loss and its impact on quality of life. Three different measures of quality of life were used: EQ-5D[12]: (£16,100/quality-adjusted life-year (QALY)), HUI3[13] (£8,710/QALY) and SF-6D[14] (£14,900/QALY).

These costs per QALY are well below the threshold of acceptable value for money for the NHS, appearing to indicate that the extra costs of PBT are justified. The results were robust to sensitivity analysis.

However, although the estimated risks of grade 3 or 4 hearing loss after PBT were similar to those reported in Paulino et al 2018, the rates after PRT were much higher. This may be because of improvements in radiotherapy techniques since the 1980s, when one of the studies (Schell et al 1989) on which Hirano et al 2014 relied was published. Hearing loss rates supported by modern evidence lie outside the sensitivity ranges used by Hirano et al 2014, casting doubt on the reliability of their conclusions.

Ependymoma

All adverse treatment effects
Sato et al 2017 reported a study involving 79 children with intracranial ependymoma, 41 of whom were treated with PBT and 38 with PRT. This paper is mainly concerned with the clinical effectiveness of the two treatments and includes only limited reporting of safety outcomes. Toxicity rates after PBT were 3/41 (7.3%), after PRT they were 5/38 (13.2%), (χ^2 = 0.237, p=0.626 with Yates' correction, calculated by SPH). Adverse effects of treatment included radiation necrosis, stroke and cavernoma.

Children receiving PBT had a median age less than half that of the PRT group, being on average 3.2 years younger. Their follow-up was also on average 2.3 years shorter, which may have biased the study in favour of PBT, as there was less time for late adverse effects to emerge.

The study does not indicate benefit from PBT.

Symptomatic MRI abnormalities
Gunther et al 2015 reported MRI abnormalities with associated symptoms in 72 children with ependymoma. In those receiving PBT, 4/37 (11%) had abnormalities with symptoms, compared with 3/35 (8.6%) after PRT (χ^2 = 0.006, p=0.938 with Yate's correction (calculated by SPH)).

The study does not indicate benefit from PBT.

[12] A standardised instrument for measuring health status
[13] The Health Utilities Index 3, a rating scale used to measure general health status and health-related quality of life
[14] Short form 6 dimension is a measure of health utility

Craniopharyngioma

Adverse treatment effects

Bishop et al 2014's study included 52 children with craniopharyngioma. The authors reported several adverse effects of treatment, though none showed a significant difference in rates between participants receiving the two treatments:

- Vascular morbidity, including moyamoya, stroke, and vessel malformations: PBT 2/21 (10%), PRT 3/31 (10%), p=1.0
- Visual morbidity: PBT 1/21 (5%), PRT 4/31 (13%), p=0.637
- Hypothalamic obesity: PBT 4/21 (19%), PRT 9/31 (29%), p=0.523
- Endocrinopathy; PBT 16/21 (76%), PRT 24/31 (77%), p=1.0.

The median length of follow-up for participants treated with PRT was more than three times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group.

In any case, Bishop et al 2014 does not indicate safety advantages from PBT.

Salivary gland tumours

Various adverse treatment effects

Grant al 2015 published a small study of 24 children with malignant salivary gland tumours. They report rates of several local adverse effects:

- Dermatitis: PBT 7/13 (54%), PRT 6/11 (55%), p=1.0
- Dysphagia: PBT 0/13 (0%), PRT 3/11 (27%), p=0.08
- Otitis externa: PBT 1/13 (8%), PRT 2/11 (18%), p=0.58
- Mucositis: PBT 6/13 (46%), PRT 10/11 (91%), p<0.05 reported by authors, p=0.0335 calculated by SPH

There are two methodological weaknesses in Grant et al 2015. Firstly, the median length of follow-up for participants treated with PRT was more than 10 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. Secondly, the authors carried out four tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Bonferroni correction of their significance threshold of p=0.05 gives an adjusted p-value of 0.05/4=0.0125. So, none of the reported differences is statistically significant.

Retinoblastoma

Second malignancies

Sethi et al 2014 reported results in 86 children with retinoblastoma, of whom 55 were treated with PBT and 31 with PRT. The rate of second malignancies in the field irradiated by PBT were 0/55 (0%), 95% confidence interval (CI) not reported; after PRT the rate was 4/31 (14%), 95% CI 3% to 31% (p=0.015). Corresponding rates for second malignancies anywhere were [figure not reported]/55 (5%), 95% CI 0% to 21%, and 4/31 (13%), 95% CI 3% to 31% respectively (p=0.120).

This paper has the same analytical flaw as Bishop et al 2014 and Grant et al 2015, and is equally unreliable. The median length of follow-up for participants treated with PRT was nearly

twice that in those who received PBT (p=0.006). However, the reporting of adverse effects was a cumulative total over ten years, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group.

Tumours at several sites

Decline in intelligence quotient

Kahalley et al 2017 reported a study of 123 children with brain tumours. The 90 who received PBT had no statistically significant decline in intelligence quotient (IQ) (p= 0.130), though no absolute value for their IQ change was reported. The children who received PRT has a loss of 1.1 IQ points per year (p= 0.004). However, a comparison of the change in IQ over time between the two groups revealed no significant difference in rates of decline (p= 0.509).

These results are contradictory, but the authors provide a clear explanation and interpretation of their findings which is compatible with their data. They conclude that *"this study does not provide clear evidence that [PBT] results in clinically meaningful sparing of global IQ significantly exceeding that of modern [PRT] protocols."* They also note that *"it is difficult to ascribe clinical meaningfulness to a difference in IQ change as small as that observed in this sample."* They suggest that *"modern [PRT] protocols may be so successful at limiting exposure to healthy surrounding brain tissue that patients treated since 2002 are not experiencing the extent of neurocognitive decline reported in previous studies".*

Health-related quality of life

Yock et al 2014 analysed the health-related QoL in 120 children with brain tumours. Using the PedsQL[15] Core Module, they reported:

- mean PedsQL total core score: PBT 75.9, PRT 65.4, unadjusted p=0.002
- physical summary score PBT 78.4, PRT 68.1, unadjusted p=0.015
- psychosocial summary score: PBT 74.5, PRT 64.0, unadjusted p=0.001.

This study is affected by several biases which mean that its conclusions are not reliable:

- Since family income and quality of life may differ between people of different ethnicities, the significantly higher proportion of children treated with PBT who were white (PBT 84.2%, PRT 50.8%, P<0.001) may explain the effects reported here.
- The authors say that *"the proton cohort likely includes a larger proportion of patients from a higher socio-economic status".* This may also have caused differences in reported quality of life between the two groups.
- The authors also note that "*The more recently treated proton cohort … may have benefited from improved techniques over time in all the treatment arenas, including surgery, chemotherapy and radiation therapy, which would skew the results to favor the proton cohort."*

Most importantly, the authors report *"As the marginal error rates are of primary interest, rather than an experiment-wise rate, the data analysis … has not been adjusted for multiple comparisons."* This unorthodox approach raises the risk of differences being deemed significant when they were the result of the many comparisons being made. This is a particular problem as the authors reported 35 comparisons, by tumour site, by subdomains of PedsQL and with a

[15] The PedsQL is a validated assessment of health-related QoL for children with or without chronic health conditions. Scores are from 0 to 100, with 100 representing the best quality of life. PedsQL total scores are in two major sub-domains, physical and psychosocial.

normative population. We therefore calculated Bonferroni-corrected p-values, dividing the standard significance level of 0.05 by the number of QoL comparisons in the paper (35) to yield an adjusted p-value of 0.00143. Using this threshold, the results for the overall comparison and for physical summary scores are not significant, while those for psychosocial summary scores are of borderline significant depending on rounding.

For these reasons, no conclusions in favour of PBT can be drawn from Yock et al 2014.

Various adverse treatment effects

Song et al 2014 reported a study of 43 children with malignancies at various sites, with measures of rates of these adverse effects:

- Leukopaenia: grade 3[16]: PBT 14/30 (57%), PRT 6/13 (46%); grade 4[17]: PBT 2/30 (7%), PRT 4/13 (31%); p=0.069
- Anaemia: grade 3[18]: PBT 0/30 (0%), PRT 2/13 (15%), p=0.493
- Thrombocytopenia: grade 3[19]: PBT 6/30 (20%), PRT 4/13 (31%); grade 4[20]: PBT 1/30 (3%), PRT 3/13 (23%); p=0.012
- Platelet transfusion: PBT 5/30 (17%), PRT 6/13 (46%), p=0.042
- Dysphagia: PBT 14/30 (47%), PRT 2/13 (15%), p=0.086
- Diarrhoea: PBT 0/30 (0%), PRT 3/13 (23%), p=0.023.

The authors carried out 15 tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Bonferroni correction of their significance threshold of p=0.05 gives an adjusted p-value of 0.05/15=0.0033. So, none of the reported differences is statistically significant.

Does delivery of radiation by protons compared with photons reduce the risks of toxicity to key organs in children and young adults? The organs at risk are: brain (cognitive dysfunction); optic nerves (visual failure); pituitary gland (endocrine dysfunction); cranial nerves; immature skeleton (growth retardation); heart; lung; rectum; bladder; reproductive system (reduced fertility); breast; optic chiasm; cochlea; hypothalamus; hippocampus/temp lobes; brainstem; spinal cord; cauda equina; kidneys; thyroid; small bowel

We found no reliable evidence that PBT is less toxic than PRT.

By how much does the delivery of radiation by protons reduce the risks of toxicity to key organs in children and young adults compared with each of the subgroups of photon therapy?

We do not know. We found no reliable evidence that the risk of toxicity is reduced in any organ.

Can the risk reduction be quantified?

No. We found no reliable evidence that a risk reduction exists.

[16] Grade 3: <2000 – 1000/mm^3 (<2.0 – 1.0 x 10^9 /L)
[17] Grade 4: <1000/mm^3 (<1.0 x 10^9 /L)
[18] Hb 6.5 to 8 g/dl.
[19] <1.0 – 0.5 x 109 /L
[20] < 0.5 x 10^9/L (< 500/mm^3)

What is the reduction of risk for late radiation second malignancy?

We do not know. The only study relevant to this question that we found was unreliable (Sethi et al 2014).

Are there any particular characteristics of the tumour or the radiation delivery strategy that increases the risk of late toxicity?

We do not know. We found no evidence relevant to this question.

What are the cost consequences of incremental toxicity associated with photon radiotherapy compared with protons?

We do not know. We found no evidence relevant to this question.

5 Discussion

There is a substantial body of evidence comparing the adverse effects of PBT and PRT. It should therefore have been possible to draw clear conclusions, either in respect of the treatments in general or in specific tumours.

However, the evidence that we found had several limitations which prevented this:

- Some studies were simply inconclusive, for example Paulino et al 2018, Gunther et al 2015 and Sato et al 2017. This may be because PBT is no safer than PRT, or it may reflect the studies' small size and consequent lack of statistical power. Kahalley et al 2017 produced mixed and contradictory results, but the authors' clear view that their study was inconclusive makes it unwise to draw other conclusions.

- Some were subject to potential biases that could explain the findings. For example, in Eaton et al 2015, the PBT participants were on average more than two years younger than those who received PRT (6.2 years vs 8.3 years). This may have affected the susceptibility of adjacent tissue to irradiation and biased the study. There were also incorrected but potentially important biases in Yock et al 2014.

- Some studies did not analyse data correctly. For example, Bishop et al counted adverse events and compared the numbers in the two groups, despite the median length of follow-up in participants treated with PRT being more than three times that in those who received PBT. This renders the results uninterpretable. Other studies with this defect include Grant et al 2015 and Sethi et al 2014.

- Some studies were apparently conclusive, but relied on incorrect statistical techniques. For example, Song et al 2014 carried out 15 tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Correction of the authors' significance threshold gave an adjusted p-value lower than that for any of the reported differences. Other studies that failed to adjust for multiple comparisons were Grant et al 2015 and Yock et al 2014.

- Some relied on obsolete data. For example, Hirano et al 2014 used rates of hearing loss after PRT from the 1980s. More recent studies such as Paulino et al 2018 report rates outside the sensitivity range used by Hirano et al 2014, rendering their cost utility

analysis unreliable.

- The participants who received PRT in some studies were treated some time ago. For example, PRT started in 1986 in Sethi et al 2014, in 1996 in Grant et al 2015 and in Bishop et al 2014, in 1998 in Yock et al 2014 and in 2002 in Kahalley et al 2017. Photon-based treatments available nowadays may be less toxic than those reported here, for example by the use of stereotactic techniques, making the comparison no longer relevant. None of the studies we found reported a comparison with stereotactic photon radiotherapy.

Taken together, these defects and limitations mean that none of the studies provides reliable evidence of safety advantages from PBT over PRT.

6 Conclusion

There is a substantial amount of evidence comparing adverse results of PBT and PRT. However, the studies that we found were inconclusive, biased and/or incorrectly analysed. None provided reason to believe that PBT is associated with a lower risk of adverse treatment effects than PRT.

Randomised trials are needed with appropriate analysis to resolve the uncertainties still present despite the studies which were included in this review.

The lack of evidence precludes conclusions about the relative safety of PBT and PRT, about the quantification of safety advantages, about effects on second malignancies or about cost implications of different treatments.

7 Evidence Summary Table

For abbreviations see list at end of section

Proton beam therapy versus photon x-ray radiotherapy in medulloblastoma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Paulino et al 2018	P1: Controlled unrandomised study Houston, USA	84 children with medulloblastoma treated with craniospinal radiation (PBT 38, PRT 46) and cisplatin-based chemotherapy between 1997 and 2013. Male: PBT 28/38 (74%), PRT 32/46 (70%), p=0.678. Median age (years): PBT 7.6 (range 2.9 to 14.5), PRT 9.0 (range 3.0 to 18.0), p=0.262.	Mean cochlear radiation dose (Gy): PBT 3150, SD 786; PRT 3726, SD 543; P<0.0001. Mean cisplatin dose (mg/m^2): PBT 281, SD 59.5; PRT 356, SD 140; p=0.004.	Primary outcome Safety	Hearing loss of grade 3 or 4 on the SIOP Boston scale[21]	PBT: 15/75 (20%), PRT 21/91 (23%), p=0.63.	8	Direct	The authors also report three other measures of hearing loss, but none showed a significant difference in its incidence between the participants treated with PBT and PRT. Cisplatin is ototoxic (Paken et al 2016). Participants receiving PRT had both higher doses of an ototoxic drug and, according to the authors' modelling, higher radiation doses to their cochleas, but there is no reported difference in the risk of hearing loss.

[21] A hearing loss scale: grade 0= ≤20dB loss at all frequencies, grade 1 = > 20dB sensorineural hearing loss (SNHL) above 4 kHz, Grade 2 = > 20dB SNHL at 4 kHz, Grade 3 = > 20 dB SNHL at or above 2 kHz, Grade 4 = Grade 2 = >40 dB SNHL at or above 2 kHz.

Proton beam therapy versus photon x-ray radiotherapy in medulloblastoma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Eaton et al 2015	P1: Controlled unrandomised study Atlanta and Boston, USA	77 children with medulloblastoma treated with craniospinal radiation (PBT 40, treated in Boston between 2000 and 2009 as part of a trial with prospective data collection) or photon radiotherapy (PRT 37, treated in Atlanta outside a trial and with retrospective data collection). Age ≥3 years at diagnosis, <1.5 cm^2 residual disease after surgery, and no metastases seen on MRI of the brain	Median cranio-spinal irradiation (CSI) dose (Gy), PBT 23.4 (range 18 to 27), PRT 23.4 (range 18 to 26.4), p=0.681.	Primary outcome Safety	Risk of hypothyroidism, sex hormone deficiency, endocrine replacement therapy and lower height.	Multivariable analysis adjusted for gender, date of diagnosis, histology, location of radiotherapy boost, age at diagnosis, and craniospinal radiation dose: hypothyroidism OR 0.13, 95% CI 0.04 to 0.41, p<0.001. sex hormone deficiency OR 0.06, 95% CI 0.01 to 0.55, p=0.013. endocrine replacement therapy OR 0.30, 95% CI 0.09 to 0.99, p=0.047. greater height standard deviation score[22] at last follow-up: parameter estimate 0.89 (indicating greater height with PBT), 95% CI 0.24 to 1.54, p=0.008.	7	Direct	The PBT participants were on average more than two years younger than those who received PRT. This may have affected the susceptibility of adjacent tissue to irradiation and biased the study. The authors suggest that the differences that they report may be due to biases in diagnostic testing and acceptance of treatment at the two hospitals. For example, the Atlanta participants, treated with PRT, were only tested for growth hormone deficiency (GHD) when it was clinically suspected, and "testing may not have been undertaken if the patient/family actively declined the treatment prior to testing. Family willingness to undergo GH replacement may have been impacted by social factors such as cost or a fear of the potential impact on tumor recurrence or second malignancy risk. This may have artificially lowered the GHD reported, as patients may have had clinical evidence of GHD but may not have undergone the confirmatory testing required to make the diagnosis." By contrast, all Boston participants were recommended to have growth hormone stimulation testing (a confirmatory test) when there was clinical suspicion of the diagnosis. Differences in the timing and purpose of data collection may also have introduced bias. The spouse of one of the authors is on the medical advisory board of a company with provided PBT, and holds stock options in the company.

[22] An indication of how close to normal height is.

Proton beam therapy versus photon x-ray radiotherapy in medulloblastoma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
		and spine and cerebrospinal fluid cytology examination. All participants had ≥3 years of follow-up with routine endocrine screening and without disease progression or receipt of salvage therapy. Male: PBT 21/40 (53%), PRT 24/37 (65%), p=0.271. Median age at diagnosis (years): PBT 6.2 (range 3.3 to 22), PRT 8.3 (range 3.4 to 19.5), p=0.01.							

Proton beam therapy versus photon x-ray radiotherapy in medulloblastoma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Hirano et al 2014	S2: Secondary analysis of existing data Health economic model	Markov model of PBT versus PRT for medulloblastoma in children of 6 years, considering only the risk of hearing loss and its impact on quality of life	Patients were modelled in an average-risk group (no metastases and residual disease < 1.5 cm^2) (70%, who received 23.4 Gy) and a high-risk group (metastatic disease or residual disease ≥ 1.5cm^2), (30%, who received 30.6 Gy). Participants also were modelled to receive a total dose of vincristine and cisplatin of "~ 300 mg". Costs were from the Japanese healthcare system in 2012.	Primary outcome Cost utility	Cost per QALY	Three different measures of quality of life were used: EQ-5D[23]: £16,100/QALY, HUI3[24]: £8,710/QALY and SF-6D[25] £14,900/QALY.	5	Direct	The 5-year survival rates were defined as 85% (95% CI 75% to 94%) for the average-risk group and 70% (95% CI 54% to 84%) for the high-risk group. Mortality after 5 years reflected general mortality rates. Costs in Japan may differ from those in the NHS. All three metrics of quality of life give estimates of cost utility well below the threshold of acceptable value of money for the NHS. The results were robust to sensitivity analysis. The estimated risks of grade 3 or 4 hearing loss were: PRT average-risk 39% (sensitivity range 37% to 41%), PRT high-risk 47.1% (sensitivity range 44.6% to 49.7%), PBT average-risk 15.6% (sensitivity range 4.97% to 26.1%), PBT high-risk 26.5% (sensitivity range 18.4% to 34.7%). The PBT rates are similar to those reported in Paulino et al 2018, but the PRT rates are much higher. This may be because of improvements in radiotherapy techniques since the 1980s, when one of the studies (Schell et al 1989) on which Hirano et al 2014 relied was published. Hearing loss rates supported by modern evidence lie outside the sensitivity ranges used by Hirano et al 2014, casting doubt on the reliability of their conclusions.

[23] A standardised instrument for measuring health status
[24] The Health Utilities Index 3, a rating scale used to measure general health status and health-related quality of life
[25] Short form 6 dimension is a measure of health utility

Proton beam therapy versus photon x-ray radiotherapy in ependymoma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Sato et al 2017	P1: Controlled unrandomised study 2 hospitals in Houston, US	79 children with intracranial ependymoma, treated with PBT (41, treated between 2006 and 2013) or PRT (38, 21 treated before 2006 and 17 treated between 2006 and 2013). Male: PBT 25/41 (61%), PRT 21/38 (55%), p=0.607. Median age at diagnosis (years): PBT 2.5 (range 0.5 to 18.7), PRT 5.7 (range 0.4 to 16.5), p=0.001. Median follow-up: PBT 2.6 years, PRT 4.9 years, P<0.0001	PBT: median dose 55.9 Gy, range 50.4 Gy to 59.40 Gy PRT: median dose 54.0 Gy, range 50.4 Gy to 59.4 Gy p= 0.056	Secondary outcome Safety	Toxicity	Toxicity rates: PBT 3/41 (7.3%), PRT 5/38 (13.2%), χ^2 = 0.237, p=0.626 with Yates' correction (calculated by SPH).	6	Direct	This paper is mainly concerned with the clinical effectiveness of the two treatments and includes only limited reporting of safety outcomes. Three children treated with PBT developed radiation necrosis, 2 in the 4th ventricle and 1 in the temporal lobe. Of the 5 adverse reactions to PRT, 3 children developed radiation necrosis (2 in the 4th ventricle and 1 in the frontoparietal region), 1 had a stroke and 1 developed a cavernoma. Children receiving PBT had a median age less than half that of the PRT group, being on average 3.2 years younger. Their follow-up was also on average 2.3 years shorter, which may have biased the study in favour of PBT, as there was less time for late adverse effects to emerge. The study had little power to detect differences in symptomatic adverse effects of treatment.

Proton beam therapy versus photon x-ray radiotherapy in ependymoma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Gunther et al 2015	P1: Controlled unrandomised study 2 hospitals in Houston, US	72 children with non-metastatic intracranial ependymoma, treated between 2000 and 2013 with PBT (37) or PRT (35). Male: PBT 22/37 (59%), PRT 19/35 (54%), p=0.45. Mean age at treatment (years): PBT 4.4 (range 1.3 to 19), PRT 6.9 (range 1.6 to 16.6), p=0.2.	Median dose: PBT 59.4 Gy, PRT 54.0 Gy, p= 0.44	Secondary outcome Safety	MRI abnormalities with associated symptoms. Reported symptoms after radiotherapy included hemiplegia, ataxia, seizures and dysarthria.	PBT 4/37 (11%), PRT 3/35 (8.6%), $\chi^2 = 0.006$, p=0.938 with Yate's correction (calculated by SPH).	7	Direct	Reported asymptomatic radiological abnormalities were out-of-scope. Patients with radiological abnormalities (mostly asymptomatic) were younger than those without (median age at treatment 2.7 years versus 4.2 years, p=0.2). Because the PBT patients were also on average younger, it is difficult to determine whether any reported differences between the two treatment groups are valid, or the result of confounding by age. The study had little power to detect differences in symptomatic adverse effects of treatment.

Proton beam therapy versus photon x-ray radiotherapy in craniopharyngioma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Bishop et al 2014	P1: Controlled unrandomised study 2 hospitals in Houston, US	52 children with craniopharyngioma, treated between 1996 and 2012 with PBT (21) or PRT (31). Male: PBT 9/21 (43%), PRT 14/31 (45%), p=1.0. Median age at treatment (years): PBT 9.1, PRT 8.8, p=1.0. Median follow-up (years): PBT 2.76, PRT 8.84, P<0.001.	Median dose: PBT 50.4 Gy, PRT 50.4 Gy.	Primary outcome Safety	Vascular morbidity, including moyamoya, stroke, and vessel malformations	PBT 2/21 (10%), PRT 3/31 (10%), p=1.0	7	Direct	The median length of follow-up for participants treated with PRT was more than 3 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group.
				Primary outcome Safety	Visual morbidity	PBT 1/21 (5%), PRT 4/31 (13%), p=0.637			
				Primary outcome Safety	Hypothalamic obesity	PBT 4/21 (19%), PRT 9/31 (29%), p=0.523			
				Primary outcome Safety	Endocrinopathy	PBT 16/21 (76%), PRT 24/31 (77%), p=1.0			

Proton beam therapy versus photon x-ray radiotherapy in salivary gland tumours									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Grant al 2015	P1: Controlled unrandomised study A hospital in Houston, US	24 children with malignant salivary gland tumours, treated between 1996 and 2014 with PBT (13) or PRT (11). Male: PBT 6/13 (46%), PRT 5/11 (45%), p=1.0. Median age at treatment (years): PBT 13, PRT 15, p=0.41. Median follow-up (years): PBT 0.67, PRT 7.7, P<0.05.	Median dose: PBT 60 Gy, PRT 60 Gy.	Primary outcome Safety	Dermatitis (brisk erythema, moderate oedema, or moist desquamation)	PBT 7/13 (54%), PRT 6/11 (55%), p=1.0	5 (2,1,1,0,1)	Direct	The median length of follow-up for participants treated with PRT was more than 10 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. The authors carried out 4 tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Bonferroni correction of their significance threshold of p=0.05 gives an adjusted p-value of 0.05/4=0.0125. So, none of the reported differences is statistically significant.
				Primary outcome Safety	Dysphagia (pain requiring change in diet and/or nutritional support)	PBT 0/13 (0%), PRT 3/11 (27%), p=0.08			
				Primary outcome Safety	Otitis externa (discharge from ear canal)	PBT 1/13 (8%), PRT 2/11 (18%), p=0.58			
				Primary outcome Safety	Mucositis (patchy or confluent ulcerations)	PBT 6/13 (46%), PRT 10/11 (91%), P<0.05 reported by authors, p=0.0335 calculated by SPH			

Proton beam therapy versus photon x-ray radiotherapy in retinoblastoma									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Sethi et al 2014	P1: Controlled unrandomised study A hospital in Houston, US	86 children with retinoblastoma, treated between 1986 and 2011 with PBT (55) or PRT (31). Male: PBT 22/55 (44%), PRT 17/31 (55%), p=0.372. Median age at treatment (months): PBT 14.8, PRT 10.0, p=0.026. Median follow-up (years): PBT 6.9, PRT 13.1, p=0.006.	Median dose: PBT 44.2 Gy, PRT 55.0 Gy, p=0.41.	Primary outcome Safety	Second malignancies in the radiation field, 10-year incidence	PBT 0/55 (0%), 95% CI not reported; PRT 4/31 (14%), 95% CI 3% to 31%; p=0.015	5	Direct	In-field location of malignancies was assumed for all tumours in the brain, orbits, facial sinuses, temporal bones or soft tissue overlying the temporal bones. The median length of follow-up for participants treated with PRT was nearly twice that in those who received PBT. However, the reporting of adverse effects was a cumulative total over ten years, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. PBT participants were also older when treated. The authors report 4/31 as 14%, whereas it is in fact 12.9%. We have corrected it to 13% here.
				Primary outcome Safety	Second malignancies anywhere, 10-year incidence	PBT [figure not reported]/55 (5%), 95% CI 0% to 21%; PRT 4/31 (13%), 95% CI 3% to 31%; p=0.120.			

Proton beam therapy versus photon x-ray radiotherapy in tumours of several primary sites									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
Kahalley et al 2017	P1: Controlled unrandomised study A hospital in Houston, US	150 children with brain tumours, treated with PBT between 2007 and 2012 (90) or with PRT between 2002 and 2007 (60). Male: PBT 54/90 (60%), PRT 33/60 (55%), p=0.543. Mean age at treatment (years): PBT 9.2, PRT 8.1, p=0.108. Follow-up not reported. Cranio-spinal irradiation PBT 51/90 (57%), PRT 31/60 (52%), p=0.547. Glioma 23 (15%, medulloblastoma 64	Median dose: PBT 54.0 Gy, PRT 54.0 Gy, p=0.01.	Primary outcome Safety	Decline in intelligence quotient, all participants	*Analysis of 123 participants* PBT: "no statistically significant decline", absolute value not reported, 95% CI for gradient -1.6 to 0.2, p= 0.130; PRT: loss of 1.1 IQ points per year, 95% CI -1.8 to -0.4; p= 0.004. Comparison of the change in IQ over time between the PBT and PRT groups: -0.7 v -1.1 points per year respectively, p= 0.509.	8	Direct	27 participants were excluded from the multivariate analysis of effects of treatment on IQ because of missing co-variate data. The authors conclude that "this study does not provide clear evidence that [PBT] results in clinically meaningful sparing of global IQ significantly exceeding that of modern [PRT] protocols." They also note that "it is difficult to ascribe clinical meaningfulness to a difference in IQ change as small as that observed in this sample." They suggest that "modern [PRT] protocols may be so successful at limiting exposure to healthy surrounding brain tissue that patients treated since 2002 are not experiencing the extent of neurocognitive decline reported in previous studies".
				Primary outcome Safety	Decline in intelligence quotient, participants who received cranio-spinal irradiation	*Analysis of 69 participants* PBT: "no statistically significant decline", absolute value and 95% CIs not reported, p=0.203; PRT: "no statistically significant decline", absolute value and 95% CIs not reported, p=0.060. Comparison of the change in IQ over time between the PBT and PRT groups: -0.8 v -0.9 points per year respectively, p=0.890.			

Proton beam therapy versus photon x-ray radiotherapy in tumours of several primary sites									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
		(43%), ependymoma 17 (11%), germ cell tumours 20 (13%), other 23 (15%).		Primary outcome Safety	Decline in intelligence quotient, participants who received focal irradiation	*Analysis of 54 participants* PBT: "no statistically significant decline", absolute value not reported, 95% CI -2.0 to 0.8, p=0.401; PRT: loss of 1.6 points per year, 95% CI -3.0 to -0.2, p= 0.026. Comparison of the change in IQ over time between the PBT and PRT groups: -0.6 v -1.6 points per year respectively, p= 0.342.			
Yock et al 2014	P1: Controlled unrandomised study 2 hospitals in Houston and in Palo Alto, both in the USA	120 children with brain tumours, treated in Houston with PBT mostly between 2006 and 2007 (57) or treated in Palo Alto with PRT mostly between 1998 and 2002 (63). Male: PBT	Median dose: PBT 50 to 54 Gy, PRT 50 to 54 Gy, no tests of significance reported.	Primary outcome Safety	Health-related QoL assessed with the parent-proxy report versions of the PedsQL[26] Core Module. The PBT cohort were assessed during treatment and then annually thereafter. The PRT cohort were assessed once, 1 to 2 years after treatment.	Mean PedsQL total core score score: PBT 75.9, PRT 65.4, unadjusted p=0.002, not significant. Physical summary score PBT 78.4, PRT 68.1, unadjusted p=0.015, not significant. Psychosocial summary score: PBT 74.5, PRT 64.0, unadjusted p=0.001, borderline significant depending on rounding.	4	Direct	Since family income and quality of life may differ between people of different ethnicities, the significantly higher proportion of children treated with PBT who were white may explain the effects reported here. The authors say that "the proton cohort likely includes a larger proportion of patients from a higher socio-economic status". They also note that "The more recently treated proton cohort … may have benefited from improved techniques over time in all the treatment arenas, including surgery, chemotherapy and radiation therapy, which would skew the results to favor the proton cohort." The authors report "As the marginal error rates are of primary interest, rather than an experiment-wise rate, the data analysis … has not been adjusted for multiple comparisons." This unorthodox approach raises the risk of differences being deemed

[26] The PedsQL is a validated assessment of health-related QoL for children with or without chronic health conditions. Scores are from 0 to 100, with 100 representing the best quality of life. PedsQL total scores are in two major sub-domains, physical and psychosocial. The psychosocial summary score is further sub divided into 3 parts: emotional functioning, social functioning and school functioning.

Proton beam therapy versus photon x-ray radiotherapy in tumours of several primary sites									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
		50.9%, PRT 55.6%, p=0.608. Mean age at treatment (years): PBT 7.0, PRT 7.7, p=0.585. Follow-up (years): PBT not reported, PRT 2.9. White ethnicity: PBT 84.2%, PRT 50.8%, P<0.001. Glioma 18 (15%), medulloblastoma 48 (40%), ependymoma 27 (23%), germ cell tumours 14 (12%), other 13 (11%).							significant when they were the result of the many comparisons being made. We have therefore calculated Bonferroni-corrected P-values, dividing the standard significance level of 0.05 by the number of QoL comparisons in the paper (35) to yield an adjusted p-value of 0.00143. Parents' scores may be less valid because of lack of first-hand knowledge of the benefits and adverse effects of treatment.
Song et al 2014	P1: Controlled unrandomised study	43 children with malignant tumours, treated with	Mean dose: PBT 51.8 Gy, PRT 53.2 Gy, p=0.858.	Primary outcome Safety	Leukopaenia	Grade 3[27]: PBT 14/30 (57%), PRT 6/13 (46%); Grade 4[28]: PBT 2/30 (7%), PRT 4/13 (31%); p=0.069	6	Direct	The authors carried out 15 tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Bonferroni correction of their significance threshold of p=0.05 gives an adjusted p-value of 0.05/15=0.0033. So, none of the

[27] Grade 3: <2000 – 1000/mm^3 (<2.0 – 1.0 x 10^9 /L)
[28] Grade 4: <1000/mm^3 (<1.0 x 10^9 /L)

Proton beam therapy versus photon x-ray radiotherapy in tumours of several primary sites									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
	A hospital in Seoul, South Korea	craniospinal irradiation with PBT (30) or with PRT (13), between 2008 and 2012. Male: PBT 16/30 (53%), PRT 8/13 62%, p=0.62. Mean age at treatment (years): PBT 10, PRT 11, p=0.25. Mean follow-up (months): 22 (range 2 to 118), not reported by treatment group. Medulloblastoma 13 (30%), germ cell tumours 19 (44%), other 8 (19%). 3 participants'			Anaemia	Grade 3[29]: PBT 0/30 (0%), PRT 2/13 (15%), p=0.493			reported differences are statistically significant.
					Thrombocytopenia	Grade 3[30]: PBT 6/30 (20%), PRT 4/13 (31%); Grade 4[31]: PBT 1/30 (3%), PRT 3/13 (23%); p=0.012			
					Platelet transfusion	PBT 5/30 (17%), PRT 6/13 (46%), p=0.042			
					Dysphagia	PBT 14/30 (47%), PRT 2/13 (15%), p=0.086			
					Diarrhoea	PBT 0/30 (0%), PRT 3/13 (23%), p=0.023.			

[29] Hb 6.5 to 8 g/dl.
[30] <1.0 – 0.5 x 109 /L
[31] < 0.5 x 10^9/L (< 500/mm^3)

Proton beam therapy versus photon x-ray radiotherapy in tumours of several primary sites									
Study reference	**Study Design**	**Population characteristics**	**Intervention**	**Outcome measure type**	**Outcome measures**	**Results**	**Quality of Evidence Score**	**Applicability**	**Critical Appraisal Summary**
		histology was omitted from the data. All the patients were seen by a radiation oncologist once a week during treatment. The first follow-up visit was one month after completing radiotherapy and then two months later.							

CI = confidence interval, HR = hazard ratio, ICER = incremental cost-effectiveness ratio, OR = odds ratio, PBT = proton beam therapy, PRT = photon radiotherapy, QALY = quality-adjusted life year, QoL = quality of life, SD = standard deviation.

8 Grade of Evidence Table

For abbreviations see list at end of section

Proton beam therapy versus photon x-ray radiotherapy in medulloblastoma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Hearing loss of grade 3 or 4 on the SIOP Boston scale[32]	Paulino et al 2018	8	Direct	B	Hearing loss of grade 3 means more than 20 dB sensorineural hearing loss at or above 2 kHz; grade 4 mean more than 40 dB sensorineural hearing loss at or above 2 kHz. Paulino et al 2018 reported that the prevalence of hearing loss of grade 3 or 4 did not differ significantly between children who had PBT and PRT. Rates of hearing loss are an objective measure of otological damage. However, they do not indicate the impact of hearing loss on the ability to carry out normal activities or on quality of life. This result indicates that there is no clinical benefit on hearing preservation from the use of PBT rather than PRT. Although the study was unrandomized, the biases were in favour of PBT, so the result is reliable.
Risk of sex hormone deficiency	Eaton et al 2015	7	Direct	B	Sex hormone deficiency was defined as a clinical diagnosis and/or initiation of treatment. Eaton et al reported a multivariable odds ratio, adjusted for gender, date of diagnosis, histology, location of radiotherapy boost, age at diagnosis and craniospinal radiation dose, of 0.06, 95% CI 0.01 to 0.55, p=0.013. An absence of sex hormone deficiency is more likely to be associated with normal sexual development. However, the authors do not report whether participants differed in rates of symptoms or in quality of life. This result is not clear or reliable. The PBT participants were on average more than two years younger than those who received PRT. This may have affected the susceptibility of adjacent tissue to irradiation and biased the study, for which multivariate analysis may not have fully adjusted, or affected the diagnostic rates of sex hormone deficiency. The authors suggest that the differences that they report may be due to biases in diagnostic testing and acceptance of treatment at the two hospitals. Differences in the timing and purpose of data collection may also have introduced bias. It is uncertain whether the reported differences would have a material impact on participants' symptoms and quality of life. The result's reliability is undermined by the non-randomised nature of the study and the differences between the two groups of participants.
Risk of hypothyroidism	Eaton et al 2015	7	Direct	B	Hypothyroidism was defined as a clinical diagnosis and/or initiation of treatment.

[32] A hearing loss scale: grade 0= ≤20dB loss at all frequencies, grade 1 = > 20dB sensorineural hearing loss (SNHL) above 4 kHz, Grade 2 = > 20dB SNHL at 4 kHz, Grade 3 = > 20 dB SNHL at or above 2 kHz, Grade 4 = Grade 2 = >40 dB SNHL at or above 2 kHz.

Proton beam therapy versus photon x-ray radiotherapy in medulloblastoma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					Eaton et al reported a multivariable odds ratio, adjusted for gender, date of diagnosis, histology, location of radiotherapy boost, age at diagnosis and craniospinal radiation dose, of 0.13, 95% CI 0.04 to 0.41, p<0.001. Normal thyroid function is more likely to be associated with normal health and development. However, the authors do not report whether participants differed in rates of symptoms or in quality of life. This result is not clear or reliable. The PBT participants were on average more than two years younger than those who received PRT. This may have affected the susceptibility of adjacent tissue to irradiation and biased the study, for which multivariate analysis may not have fully adjusted. The authors suggest that the differences that they report may be due to biases in diagnostic testing and acceptance of treatment at the two hospitals. Differences in the timing and purpose of data collection may also have introduced bias. It is uncertain whether the reported differences would have a material impact on participants' symptoms and quality of life. The result's reliability is undermined by the non-randomised nature of the study and the differences between the two groups of participants.
Risk of endocrine replacement therapy	Eaton et al 2015	7	Direct	B	Endocrine replacement therapy was defined as the initiation of treatment for an endocrine abnormality. Eaton et al reported a multivariable odds ratio, adjusted for gender, date of diagnosis, histology, location of radiotherapy boost, age at diagnosis and craniospinal radiation dose, of 0.30, 95% CI 0.09 to 0.99, p=0.047. An absence of endocrine deficiency is more likely to be associated with normal health and development. However, the authors do not report whether participants differed in rates of symptoms or in quality of life. This result is not clear or reliable. The PBT participants were on average more than two years younger than those who received PRT. This may have affected the susceptibility of adjacent tissue to irradiation and biased the study, for which multivariate analysis may not have fully adjusted. The authors suggest that the differences that they report may be due to biases in diagnostic testing and acceptance of treatment at the two hospitals. Differences in the timing and purpose of data collection may also have introduced bias. It is uncertain whether the reported differences would have a material impact on participants' symptoms and quality of life. The result's reliability is undermined by the non-randomised nature of the study and the differences between the two groups of participants.
Risk of lower height	Eaton et al 2015	7	Direct	B	Lower height was defined as having a lower standard deviation score, a measure of difference in height. Eaton et al reported a parameter score, adjusted for gender, date of diagnosis, histology, location of radiotherapy boost, age at diagnosis and craniospinal radiation dose, of 0.89 (indicating greater height with PBT), 95% CI 0.24 to

Proton beam therapy versus photon x-ray radiotherapy in medulloblastoma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					1.54, p=0.008. Less than normal height is not desirable. However, the authors do not report how much shorter PRT participants were in absolute terms and whether this affected their quality of life. This result is not clear or reliable. The PBT participants were on average more than two years younger than those who received PRT. This may have affected the susceptibility of adjacent tissue to irradiation and biased the study, for which multivariate analysis may not have fully adjusted. The authors suggest that the differences that they report may be due to biases in diagnostic testing and acceptance of treatment at the two hospitals. Differences in the timing and purpose of data collection may also have introduced bias. It is uncertain whether the reported differences would have a material impact on participants' symptoms and quality of life. The result's reliability is undermined by the non-randomised nature of the study and the differences between the two groups of participants.
Cost per QALY	Hirano et al 2014	5	Direct	C	Cost per QALY is the incremental cost of one treatment over a less expensive one, divided by the extra QALYs which it yields. Hirano et al 2014 reported three different measures of quality of life, with these costs per QALY: EQ-5D[33] £16,100, HUI3[34] £8,710 and SF-6D[35] £14,900. A lower incremental cost effectiveness ratio indicates better value for money. This does not directly benefit individual patients, but means that more patients can be treated with the resources available. All three metrics of quality of life give estimates of cost utility well below the threshold of acceptable value of money for the NHS. The results were robust to sensitivity analysis, but costs in Japan may differ from those in the NHS. The estimated risks of grade 3 or 4 hearing loss were: PRT average-risk 39% (sensitivity range 37% to 41%), PRT high-risk 47.1% (sensitivity range 44.6% to 49.7%), PBT average-risk 15.6% (sensitivity range 4.97% to 26.1%), PBT high-risk 26.5% (sensitivity range 18.4% to 34.7%). The PBT rates are similar to those reported in Paulino et al 2018, but the PRT rates are much higher. This may be because of improvements in radiotherapy techniques since the 1980s, when one of the studies (Schell MJ et al 1989) on which Hirano et al 2014 relied was published. So, hearing loss rates supported by modern evidence lie outside the sensitivity ranges used by Hirano et al 2014, casting doubt on the reliability of their conclusions.

[33] A standardised instrument for measuring health status
[34] The Health Utilities Index 3, a rating scale used to measure general health status and health-related quality of life
[35] Short form 6 dimension is a measure of health utility

Proton beam therapy versus photon x-ray radiotherapy in ependymoma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Toxicity	Sato et al 2017	6	Direct	C	Sato et al 2017 defined toxicity as any adverse reaction to treatment. Toxicity rates after PBT were 3/41 (7.3%), and after PRT 5/38 (13.2%), χ^2 = 0.237, p=0.626 with Yates' correction (calculated by SPH). Three children treated with PBT developed radiation necrosis, 2 in the 4th ventricle and 1 in the temporal lobe. Of the 5 adverse reactions to PRT, 3 children developed radiation necrosis (2 in the 4th ventricle and 1 in the frontoparietal region), 1 had a stroke and 1 developed a cavernoma. The avoidance of adverse treatment effects is valuable to patients, but Sato et al 2017 do not report the effect of these on symptoms or quality of life. This study does not indicate a difference between PBT and PRT in rates of adverse treatment effects. Children receiving PBT had a median age less than half that of the PRT group, being on average 3.2 years younger. Their follow-up was also on average 2.3 years shorter, which may have biased the study in favour of PBT, as there was less time for late adverse effects to emerge. The study had little power to detect differences in symptomatic adverse effects of treatment, reducing its reliability.
MRI abnormalities with associated symptoms	Gunther et al 2015	7	Direct	B	Gunther et al 2015 reported participants who had both an abnormality seen on MRI and an associated symptom. Reported asymptomatic radiological abnormalities were out-of-scope. Reported symptoms after radiotherapy included hemiplegia, ataxia, seizures and dysarthria. Rates of MRI abnormalities with associated symptoms after PBT were 4/37 (11%) and after PRT 3/35 (8.6%), χ^2 = 0.006, p=0.938 with Yate's correction (calculated by SPH). Reductions in rates of symptomatic adverse treatment events would benefit patients. This study does not indicate that PRT is any safer than PRT. Patients with radiological abnormalities (mostly asymptomatic) were younger than those without (median age at treatment 2.7 years versus 4.2 years, p=0.2). Because the PBT patients were also on average younger, it is difficult to determine whether any reported differences between the two treatment groups are valid, or the result of confounding by age. The study had little power to detect differences in symptomatic adverse effects of treatment, reducing its reliability.

Proton beam therapy versus photon x-ray radiotherapy in craniopharyngioma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Vascular morbidity	Bishop et al 2014	7	Direct	B	Vascular morbidity included moyamoya, stroke, and vessel malformations. Rates of vascular morbidity after PBT were 2/21 (10%), and after PRT were 3/31 (10%), p=1.0. Reductions in rates of symptomatic adverse treatment events would benefit patients. This study does not indicate that PRT is any safer than PRT. The median length of follow-up for participants treated with PRT was more than 3 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group, which limits the study's reliability.
Visual morbidity	Bishop et al 2014	7	Direct	B	Visual morbidity included any deviation in baseline vision (field cuts or acuity) on physical and ophthalmologic examination. Rates of visual morbidity after PBT were 1/21 (5%), and after PRT were 4/31 (13%), p=0.637. Reductions in rates of symptomatic adverse treatment events would benefit patients. This study does not indicate that PRT is any safer than PRT. The median length of follow-up for participants treated with PRT was more than 3 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group, which limits the study's reliability.
Hypothalamic obesity	Bishop et al 2014	7	Direct	B	Hypothalamic obesity was diagnosed "on the primary clinician's diagnosis of morbid or hypothalamic obesity during follow-up". Rates of hypothalamic obesity after PBT were 4/21 (19%), and after PRT were 9/31 (29%), p=0.523 Reductions in rates of symptomatic adverse treatment events would benefit patients. This study does not indicate that PRT is any safer than PRT. The median length of follow-up for participants treated with PRT was more than 3 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group, which limits the study's reliability.
Endocrinopathy	Bishop et al 2014	7	Direct	B	Endocrinopathy means disorders of the endocrine system; it is not defined by Bishop et al 2014.

Proton beam therapy versus photon x-ray radiotherapy in craniopharyngioma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					Rates of endocrinopathy after PBT were 16/21 (76%), and after PRT were 24/31 (77%), p=1.0 Reductions in rates of symptomatic adverse treatment events would benefit patients. This study does not indicate that PRT is any safer than PRT. The median length of follow-up for participants treated with PRT was more than 3 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group, which limits the study's reliability.

Proton beam therapy versus photon x-ray radiotherapy in salivary gland tumours					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Dermatitis	Grant al 2015	5	Direct	C	Dermatitis was defined as brisk erythema, moderate oedema or moist desquamation. Rates of dermatitis after PBT were 7/13 (54%), and after PRT were 6/11 (55%), p=1.0 The avoidance of dermatitis would be of benefit to patients. This study does not indicate that PBT is less likely to cause dermatitis than PRT. The median length of follow-up for participants treated with PRT was more than 10 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. The study was small and underpowered, so its results are less reliable.
Dysphagia	Grant al 2015	5	Direct	C	Dysphagia was defined as pain requiring change in diet and/or nutritional support. Rates of dysphagia after PBT were 0/13 (0%), and after PRT were 3/11 (27%), p=0.08. The avoidance of dysphagia would be of benefit to patients. This study does not indicate that PRT is any less likely to cause dysphagia than PRT. The median length of follow-up for participants treated with PRT was more than 10 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of

Proton beam therapy versus photon x-ray radiotherapy in salivary gland tumours					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					adverse events may have been significantly higher among the PBT group. The study was small and underpowered, so its results are less reliable.
Otitis externa	Bishop et al 2014	7	Direct	B	Otitis externa was defined by Bishop et al as discharge from ear canal. Rates of otitis externa after PBT were 1/13 (8%), and after PRT were 2/11 (18%), p=0.58. The avoidance of otitis externa would be of benefit to patients. This study does not indicate that PRT is any less likely to cause otitis externa than PRT. The median length of follow-up for participants treated with PRT was more than 3 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. The study was small and underpowered, so its results are less reliable.
Mucositis	Bishop et al 2014	7	Direct	B	Mucositis was defined by Bishop et al 2014 as patchy or confluent ulcerations. Rates of mucositis after PBT were 6/13 (46%), and after PRT were 10/11 (91%), p<0.05 reported by authors, p=0.0335 calculated by SPH The avoidance of mucositis would be of benefit to patients. This study does not indicate that PRT is any less likely to cause mucositis than PRT. The median length of follow-up for participants treated with PRT was more than 3 times that in those who received PBT. However, the reporting of adverse effects was a simple count, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. The study was small and underpowered, so its results are less reliable.

Proton beam therapy versus photon x-ray radiotherapy in retinoblastoma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Second malignancies in the radiation field, 10-year incidence	Sethi et al 2014	5	Direct	C	Second malignancies arising in the field irradiated to treat the retinoblastoma are new tumours in the brain, orbits, facial sinuses, temporal bones or soft tissue overlying the temporal bones. Rates of second malignancies after PBT were 0/55 (0%), 95% CI not reported, and after PRT were 4/31 (14%), 95% CI 3% to 31%; p=0.015. A reduced risk of secondary malignancies would be of great benefit to patients. This study does not indicate a benefit from PBT. The median length of follow-up for participants treated with PRT was nearly twice that in those who

Proton beam therapy versus photon x-ray radiotherapy in retinoblastoma					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					received PBT. However, the reporting of adverse effects was a cumulative total over ten years, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. PBT participants were also older when treated, another potential source of bias. The study was small and underpowered, so its results are less reliable.
Second malignancies anywhere, 10-year incidence	Sethi et al 2014	5	Direct	C	Second malignancies are new tumours arising anywhere in the body. Rates of second malignancies after PBT were [figure not reported]/55 (5%), 95% CI 0% to 21%, and after PRT were 4/31 (13%), 95% CI 3% to 31%; p=0.120. A reduced risk of secondary malignancies would be of great benefit to patients. This study does not indicate a benefit from PBT. The median length of follow-up for participants treated with PRT was nearly twice that in those who received PBT. However, the reporting of adverse effects was a cumulative total over ten years, not an annual rate, and the authors made no adjustment for duration of follow-up. The annual rate of adverse events may have been significantly higher among the PBT group. PBT participants were also older when treated, another potential source of bias. The study was small and underpowered, so its results are less reliable.

Proton beam therapy versus photon x-ray radiotherapy tumours of several primary sites					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Decline in intelligence quotient, all participants	Kahalley et al 2017	8	Direct	B	Intelligence quotient is an age-adjusted measure of reasoning skills. After PBT, Kahalley at al 2017 report "no statistically significant decline" in IQ, though the absolute value is not reported, 95% CI for gradient -1.6 to 0.2, p= 0.130. After PRT, there was a loss of 1.1 IQ points per year, 95% CI -1.8 to -0.4; p= 0.004. The change in IQ over time in the PBT and PRT groups was -0.7 and -1.1 points per year respectively, p= 0.509. A reduced risk of loss of intelligence would be of benefit to patients. This result does not indicate a benefit in intelligence preservation from PBT, because the results are inconsistent and indicate at best a small difference in intelligence quotients between the two treatments. The authors conclude that "this study does not provide clear evidence that [PBT] results in clinically meaningful sparing of global IQ significantly exceeding that of modern [PRT] protocols." They also note that "it is difficult to ascribe clinical meaningfulness to a difference in IQ change as small as that observed in this sample." They

Proton beam therapy versus photon x-ray radiotherapy tumours of several primary sites					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					suggest that "modern [PRT] protocols may be so successful at limiting exposure to healthy surrounding brain tissue that patients treated since 2002 are not experiencing the extent of neurocognitive decline reported in previous studies". The study's results appear reliable.
Decline in intelligence quotient, participants who received cranio-spinal irradiation	Kahalley et al 2017	8	Direct	B	Intelligence quotient is an age-adjusted measure of reasoning skills. After PBT, Kahalley at al 2017 report "no statistically significant decline" in intelligence quotient, though the absolute value and 95% CI for gradient are not reported. After PRT, there was "no statistically significant decline", absolute value and 95% CIs not reported, p=0.060. The change in IQ over time in the PBT and PRT groups was -0.8 and -0.9 points per year respectively, p= 0.890. A reduced risk of loss of intelligence would be of benefit to patients. This result does not indicate a benefit in intelligence preservation from PBT participants who received cranio-spinal irradiation. The result appears reliable.
Decline in intelligence quotient, participants who received focal irradiation	Kahalley et al 2017	8	Direct	B	Intelligence quotient is an age-adjusted measure of reasoning skills. After PBT, Kahalley at al 2017 report "no statistically significant decline" in IQ, though the absolute value is not reported, 95% CI for gradient 95% CI -2.0 to 0.8, p=0.401. After PRT, there was a loss of 1.6 points per year, 95% CI -3.0 to -0.2, p= 0.026. The change in IQ over time in the PBT and PRT groups was -0.6 and -1.6 points per year respectively, p= 0.342. A reduced risk of loss of intelligence would be of benefit to patients. This result does not indicate a benefit in intelligence preservation from PBT in participants who received focal irradiation, because the results are inconsistent and indicate at best a small difference in intelligence quotient. The authors conclude that "this study does not provide clear evidence that [PBT] results in clinically meaningful sparing of global IQ significantly exceeding that of modern [PRT] protocols." They also note that "it is difficult to ascribe clinical meaningfulness to a difference in IQ change as small as that observed in this sample." They suggest that "modern [PRT] protocols may be so successful at limiting exposure to healthy surrounding brain tissue that patients treated since 2002 are not experiencing the extent of neurocognitive decline reported in previous studies". The study's results appear reliable.

Proton beam therapy versus photon x-ray radiotherapy tumours of several primary sites					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Health-related QoL	Yock et al 2014	4	Direct	C	Health-related QoL was assessed with the parent-proxy report versions of the PedsQL[36] Core Module. It assesses QoL in two domains: physical (concerned with active daily living) and psychosocial (concerned with mood and interpersonal relationships). Yock et al 2014 report mean PedsQL total core scores of 75.9 after PBT, and 65.4 after PRT, unadjusted p=0.002, not significant. An improvement in QoL of meaningful size would be of great benefit to patients. However, this result is neither reliable nor statistically significant. Since family income and quality of life may differ between people of different ethnicities, the significantly higher proportion of children treated with PBT who were white may explain the effects reported here. The authors say that "The proton cohort likely includes a larger proportion of patients from a higher socio-economic status". They also note that "The more recently treated proton cohort … may have benefited from improved techniques over time in all the treatment arenas, including surgery, chemotherapy and radiation therapy, which would skew the results to favor the proton cohort." The authors report "As the marginal error rates are of primary interest, rather than an experiment-wise rate, the data analysis … has not been adjusted for multiple comparisons." This unorthodox approach raises the risk of differences being deemed significant when they were the result of the many comparisons being made. We have therefore calculated Bonferroni-corrected P-values, dividing the standard significance level of 0.05 by the number of QoL comparisons in the paper (35) to yield an adjusted p-value of 0.00143. Parents' scores may be less valid because of lack of first-hand knowledge of the benefits and adverse effects of treatment.
Health-related QoL, physical summary score	Yock et al 2014	4	Direct	C	Health-related QoL was assessed with the parent-proxy report versions of the PedsQL[37] Core Module. The physical summary score is concerned with active daily living. Yock et al 2014 report mean PedsQL physical summary scores of 78.4 after PBT, and 68.1 after PRT, unadjusted p=0.015, not significant. An improvement in physical summary score QoL of meaningful size would be of great benefit to patients. However, this result is neither reliable nor statistically significant. Since family income and quality of life may differ between people of different ethnicities, the

[36] The PedsQL is a validated assessment of HRQoL for children with or without chronic health conditions. Scores are from 0 to 100, with 100 representing the best quality of life. PedsQL total scores are in two major sub-domains, physical and psychosocial. The psychosocial summary score is further sub divided into 3 parts: emotional functioning, social functioning and school functioning.

[37] The PedsQL is a validated assessment of HRQoL for children with or without chronic health conditions. Scores are from 0 to 100, with 100 representing the best quality of life. PedsQL total scores are in two major sub-domains, physical and psychosocial. The psychosocial summary score is further sub divided into 3 parts: emotional functioning, social functioning and school functioning.

Proton beam therapy versus photon x-ray radiotherapy tumours of several primary sites					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					significantly higher proportion of children treated with PBT who were white may explain the effects reported here. The authors say that "the proton cohort likely includes a larger proportion of patients from a higher socio-economic status". They also note that "The more recently treated proton cohort … may have benefited from improved techniques over time in all the treatment arenas, including surgery, chemotherapy and radiation therapy, which would skew the results to favor the proton cohort." The authors report "As the marginal error rates are of primary interest, rather than an experiment-wise rate, the data analysis … has not been adjusted for multiple comparisons." This unorthodox approach raises the risk of differences being deemed significant when they were the result of the many comparisons being made. We have therefore calculated Bonferroni-corrected P-values, dividing the standard significance level of 0.05 by the number of QoL comparisons in the paper (35) to yield an adjusted p-value of 0.00143. Parents' scores may be less valid because of lack of first-hand knowledge of the benefits and adverse effects of treatment.
Health-related QoL, psychosocial summary score	Yock et al 2014	4	Direct	C	Health-related QoL was assessed with the parent-proxy report versions of the PedsQL[38] Core Module. The psychosocial summary score is concerned with mood and interpersonal relationships. Yock et al 2014 report mean PedsQL psychosocial summary score of 74.5 after PBT, and 64.0 after PRT, unadjusted p=0.001, borderline significant depending on rounding. An improvement in psychosocial QoL of meaningful size would be of great benefit to patients. However, this result is neither reliable nor of clear statistical significance. Since family income and quality of life may differ between people of different ethnicities, the significantly higher proportion of children treated with PBT who were white may explain the effects reported here. The authors say that "the proton cohort likely includes a larger proportion of patients from a higher socio-economic status". They also note that "The more recently treated proton cohort … may have benefited from improved techniques over time in all the treatment arenas, including surgery, chemotherapy and radiation therapy, which would skew the results to favor the proton cohort." The authors report "As the marginal error rates are of primary interest, rather than an experiment-wise rate, the data analysis … has not been adjusted for multiple comparisons." This unorthodox approach raises the risk of differences being deemed significant when they were the result of the many comparisons being made. We have therefore calculated Bonferroni-corrected P-values, dividing the standard significance level of 0.05 by the number of QoL comparisons in the paper (35) to yield an adjusted p-value of 0.00143. Parents' scores may be less valid because of lack of first-hand knowledge of the benefits and adverse

[38] The PedsQL is a validated assessment of HRQoL for children with or without chronic health conditions. Scores are from 0 to 100, with 100 representing the best quality of life. PedsQL total scores are in two major sub-domains, physical and psychosocial. The psychosocial summary score is further sub divided into 3 parts: emotional functioning, social functioning and school functioning.

Proton beam therapy versus photon x-ray radiotherapy tumours of several primary sites					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
					effects of treatment.
Leukopaenia	Song et al 2014	6	Direct	C	Leukopaenia is an abnormally low level of white cells in the bloodstream. Song et al reported rates of grade 3[39] leukopaenia after PBT of 14/30 (57%), and after PRT of 6/13 (46%); rates of grade 4[40] leukopaenia were 2/30 (7%) and 4/13 (31%); p=0.069. A reduced risk of leukopaenia would be of benefit to patients if it led to a lower incidence of infection. This result does not indicate a significant reduction in the risk of leukopaenia from the use of PBT. It is based on small numbers and therefore not reliable.
Anaemia	Song et al 2014	6	Direct	C	Anaemia is an abnormally low level of haemoglobin in the bloodstream. Song et al reported rates of grade 3[41] anaemia after PBT of 0/30 (0%), and after PRT of 2/13 (15%), p=0.493. A reduced risk of anaemia would be of benefit to patients if it led to reduced symptoms. This result does not indicate a significant reduction in the risk of anaemia from the use of PBT. It is based on small numbers and therefore not reliable.
Thrombocytopaenia	Song et al 2014	6	Direct	C	Thrombocytopaenia is an abnormally low number of platelets in the bloodstream. Song et al reported rates of grade 3[42] thrombocytopaenia after PBT of 6/30 (20%), and after PRT of 4/13 (31%); rates of grade 4[43] thrombocytopaenia were 1/30 (3%) and 3/13 (23%); p=0.012. A reduced risk of thrombocytopaenia would be of benefit to patients if it led to reduced symptoms. This result does not indicate a significant reduction in the risk of thrombocytopaenia from the use of PBT. The authors carried out 15 tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Bonferroni correction of their significance threshold of p=0.05 gives an adjusted P-value of 0.05/15=0.0033. So, the reported difference is not statistically significant. It is also based on small numbers and therefore not reliable.

[39] Grade 3: <2000 – 1000/mm^3 (<2.0 – 1.0 x 10^9 /L)
[40] Grade 4: <1000/mm^3 (<1.0 x 10^9 /L)
[41] Hb 6.5 to 8 g/dl.
[42] <1.0 – 0.5 x 109 /L
[43] < 0.5 x 10^9/L (< 500/mm^3)

Proton beam therapy versus photon x-ray radiotherapy tumours of several primary sites					
Outcome Measure	**Reference**	**Quality of Evidence Score**	**Applicability**	**Grade of Evidence**	**Interpretation of Evidence**
Platelet transfusion	Song et al 2014	6	Direct	C	Platelet transfusion is a treatment of thrombocytopaenia, an abnormally low number of platelets in the bloodstream. Song et al reported rates of platelet transfusion after PBT of 5/30 (17%), and after PRT of 6/13 (46%), p=0.042. A reduced risk of thrombocytopaenia would be of benefit to patients if it led to reduced need for platelet transfusion. This result does not indicate a significant reduction in the risk of platelet transfusion from the use of PBT. The authors carried out 15 tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Bonferroni correction of their significance threshold of p=0.05 gives an adjusted P-value of 0.05/15=0.0033. So, the reported difference is not statistically significant. It is also based on small numbers and therefore not reliable.
Dysphagia	Song et al 2014	6	Direct	C	Dysphagia is difficulty with or pain on swallowing. Song et al reported rates of dysphagia after PBT of 14/30 (47%), and after PRT of 2/13 (15%), p=0.086. A reduced risk of dysphagia would be of benefit to patients. This result does not indicate a significant reduction in the risk of dysphagia from the use of PBT. It is based on small numbers and therefore not reliable.
Diarrhoea	Song et al 2014	6	Direct	C	Diarrhoea is the passage of frequent loose bowel movements. Song et al reported rates of diarrhoea after PBT of 0/30 (0%), and after PRT of 3/13 (23%), p=0.023. A reduced risk of diarrhoea would be of benefit to patients. This result does not indicate a significant reduction in the risk of diarrhoea from the use of PBT. The authors carried out 15 tests of statistical significance, but did not adjust the level of statistical significance to reflect this. Bonferroni correction of their significance threshold of p=0.05 gives an adjusted P-value of 0.05/15=0.0033. So, the reported difference is not statistically significant. It is also based on small numbers and therefore not reliable. It is also based on small numbers and therefore not reliable.

CI = confidence interval, HR = hazard ratio, ICER = incremental cost-effectiveness ratio, IQ = intelligence quotient, OR = odds ratio, PBT = proton beam therapy, PRT = photon radiotherapy, QALY = quality-adjusted life year, QoL = quality of life, SD = standard deviation.

9 Literature Search Terms

Search strategy *Indicate all terms used in the search*	
P – Patients / Population Which patients or populations of patients are we interested in? How can they be best described? Are there subgroups that need to be considered?	Children (< 16 years) OR young adults (16 to about 24 years) requiring radiotherapy AND curable disease AND no distant metastases
I – Intervention Which intervention, treatment or approach should be used?	Proton beam therapy Radiotherapy with Protons Protons Particle Therapy
C – Comparison What is/are the main alternative/s to compare with the intervention being considered?	(Photon) Radiotherapy *Subgroups* IMRT (Intensity Modulated Radiotherapy) Stereotactic radiotherapy Stereotactic ablative body radiotherapy
O – Outcomes What is really important for the patient? Which outcomes should be considered? Examples include intermediate or short-term outcomes; mortality; morbidity and quality of life; treatment complications; adverse effects; rates of relapse; late morbidity and re-admission; return to work, physical and social functioning, resource use.	Acute toxicity Morbidity Toxicity Neurocognitive function Neuropsychological effects IQ CNS effects Late radiation effects Radiation toxicity Late side effects Growth retardation Cardiac toxicity Lung radiation toxicity Thyroid function Endocrine function Fertility Radiation induced second malignancy Late second malignancy Other evaluations of Quality of Life

Assumptions / limits applied to search

- English language
- Peer reviewed publications
- Clinical outcome research
- Exclude physics planning papers such as dosimetric planning
- Exclude conference abstracts
- Publications from 2008
- The literature is likely to be found for the following tumour types to aid search strategy: Medulloblastoma; Ependymoma; Rhabdomyosarcoma; Low Grade Glioma; Ewing's; Craniopharyngioma
- If there is an evidence base of sufficient size then the evidence review could be sub-divided into the different organs affected by toxicity and late effects
- Include any cost consequence studies of incremental toxicity of photons compared with protons

10 Search Strategy

We searched PubMed, Embase and Cochrane Library limiting the search to papers published in England from 1 January 2008 and 2 March 2018. We excluded conference abstracts, commentaries, letters, editorials and case reports.

Search date: 2 March 2018

Embase search:

# ▲	Searches
1	exp Neoplasms/
2	(cancer? or neoplas* or malignan* or tumour? or tumor? or carcinoma? or sarcoma? or blastoma? or glioma? or medulloblastoma? or ependymoma? or rhabdomyosarcoma? or craniopharyngioma? or ewing*).ti,ab.
3	1 or 2
4	exp adolescent/ or exp child/ or young adult/
5	(child* or schoolchild* or preschooler? or pre-schooler? or girl? or boy? or infant? or baby or babies or adolescen* or teen* or young adult* or young people or young men or young women or young male? or young female?).ti,ab.
6	4 or 5
7	Proton Therapy/
8	((proton* or particle) adj3 (therap* or radiotherap* or treatment)).ti,ab.
9	7 or 8
10	(ae or co or de).fs.
11	(safe or safety or side effect* or undesirable effect* or treatment emergent or tolerability or toxicit* or adrs or (adverse adj2 (effect or effects or reaction or reactions or event or events or outcome or outcomes))).af.
12	((cognit* or neurocognit* or psych* or mental* or neuropsych* or brain* or cereb* or spin* or ear* or hearing or otolog* or eye? or visual? or optic*) adj5 (event? or effect? or outcome? or function* or dysfunction* or disturbance? or development?)).ti,ab.
13	(iq or intelligence or literacy or numeracy or learning or language or ((academic or education*) adj2 (attain* or achiev* or status))).ti,ab.
14	(growth* adj5 (retard* or restrict* or stunt*)).ti,ab.
15	((cardi* or heart or lung or pulmonary or breast or thyroid or endocrin* or pituitary or renal or kidney? or liver or hepat* or bladder? or rect* or bowel or colorect* or colon* or intestin*) adj5 (event? or effect? or outcome? or function* or dysfunction* or disturbance?)).ti,ab.
16	(cardiotoxic* or pulmotoxic* or pneumotoxic* or thyrotoxic* or endotoxic* or renotoxic* or ototoxic* or optotoxic*).ti,ab.
17	(fertil* or subfertil* or infertil*).mp.
18	("quality of life" or QoL or HRQoL or HR-QoL).mp.
19	exp radiation injury/
20	((future or late or second*) adj5 (cancer? or neoplas* or malignan* or tumour? or tumor? or carcinoma? or sarcoma? or blastoma?)).ti,ab.
21	((radiat* or radiotherap*) adj5 injur*).ti,ab.
22	((longterm or long-term) adj3 (outcome? or effect? or consequence? or impact?)).ti,ab.
23	10 or 11 or 12 or 13 or 14 or 15 or 17 or 18 or 19 or 20 or 21 or 22
24	3 and 6 and 9 and 23
25	(editorial or letter or note or "review" or conference*).pt. or case report/
26	24 not 25
27	3 and 6 and 9
28	limit 27 to ("reviews (maximizes specificity)" or "therapy (best balance of sensitivity and specificity)")

29 limit 27 to "economics (best balance of sensitivity and specificity)"
30 26 or 28 or 29
31 limit 30 to (english language and yr="2008 -Current")

11 Evidence Selection

- Total number of publications reviewed: 146
- Total number of publications considered potentially relevant: 22
- Total number of publications selected for inclusion in this briefing: 11

12 References

Bishop AJ, Greenfield B, Mahajan A, Paulino AC, Okcu MF, Allen PK, Chintagumpala M, Kahalley LS, McAleer MF, McGovern SL, Whitehead WE, Grosshans DR 2014. Proton beam therapy versus conformal photon radiation therapy for childhood craniopharyngioma: multi-institutional analysis of outcomes, cyst dynamics, and toxicity. *Int J Radiat Oncol Biol Phys* 90: 354-361.

Grant SR, Grosshans DR, Bilton SD, Garcia JA, Amin M, Chambers MS, McGovern SL, McAleer MF, Morrison WH, Huh WW, Kupferman ME, Mahajan A 2015. Proton versus conventional radiotherapy for pediatric salivary gland tumors: acute toxicity and dosimetric characteristics. *Radiother Oncol* 116: 309-15.

Gunther JR, Sato M, Chintagumpala M, Ketonen L, Jones JY, Allen PK, Paulino AC, Okcu MF, Su JM, Weinberg J, Boehling NS, Khatua S, Adesina A, Dauser R, Whitehead WE, Mahajan A 2015. Imaging changes in pediatric intracranial ependymoma patients treated with proton beam radiation therapy compared to intensity modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 93: 54-63.

Hirano E, Fuji H, Onoe T, Kumar V, Shirato H, Kawabuchi K 2014. Cost-effectiveness analysis of cochlear dose reduction by proton beam therapy for medulloblastoma in childhood. *J Rad Res* 55: 320-7.

Kahalley LS, Ris MD, Grosshans DR, Okcu MF, Paulino AC, Chintagumpala M, Moore BD, Guffey D, Minard CG, Stancel HH, Mahajan A 2016. Comparing intelligence quotient change after treatment with proton versus photon radiation therapy for pediatric brain tumors. *J Clin Oncol* 34: 1043-1049.

NICE. www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-interventional-procedures-guidance/ip-list

Paken L, Govender CD, Pillay M, Sewram V. Cisplatin-associated ototoxicity: a review for the health professional. *J Toxicol* 2016 doi: 10.1155/2016/1809394

Paulino AC, Mahajan A, Ye R, Grosshans DR, Fatih Okcu M, Su J, McAleer MF, McGovern S, Mangona VA, Chintagumpala M 2018. Ototoxicity and cochlear sparing in children with medulloblastoma: Proton vs. photon radiotherapy. *Radiother Oncol* doi.org/10.1016/j.radonc.2018.01.002

Sato M, Gunther JR, Mahajan A, Jo E, Paulino AC, Adesina AM, Jones JY, Ketonen LM, Su JM, Okcu MF, Khatua S, Dauser RC, Whitehead WE, Weinberg J, Chintagumpala MM 2017. Progression-free survival of children with localized ependymoma treated with intensity-modulated radiation therapy or proton-beam radiation therapy. *Cancer* 123: 2570-8.

Schell MJ, McHaney VA, Green AA, Kun LE, Hayes FA, Horowitz M, Meyer WH 1989. Hearing loss in children and young adults receiving cisplatin with or without prior cranial irradiation. *J Clin Oncol* 7: 754-60.

Sethi RV, Shih HA, Yeap BY, Mouw KW, Petersen R, Kim DY, Munzenrider JE, Grabowski E, Rodriguez-Galindo C, Yock TI, Tarbell NJ, Marcus KJ, Mukai S, MacDonald SM 2014. Second non-ocular tumors among survivors of retinoblastoma treated with contemporary photon and proton radiotherapy. *Cancer* 120: 126-33.

Song S, Park HJ, Yoon JH, Kim DW, Park J, Shin D, Shin SH, Kang HJ, Kim S-K, Phi JH, Kim J-Y. Proton beam therapy reduces the incidence of acute haematological and gastrointestinal toxicities associated with craniospinal irradiation in pediatric brain tumors. *Acta Oncologica* 53: 1158-1164.

Yock TI, Bhat S, Szymonifka J, Yeap BY, Delahaye J, Donaldson SS, MacDonald SM, Pulsifer MB, Hill KS, DeLaney TF, Ebb D, Huang M, Tarbell NJ, Fisher PG, Kuhlthau KA 2014. Quality of life outcomes in proton and photon treated pediatric brain tumor survivors. *Radiother Oncol* 113: 89-94.

Zacharia BE, Bruce SS, Goldstein H, Malone HR, Neugut AI, Bruce JN 2012. Incidence, treatment and survival of patients with craniopharyngioma in the surveillance, epidemiology and end results program. *Neuro Oncol* 14: 1070-1078.