Pentosan polysulfate sodium for treating bladder pain syndrome: A Single Technology Appraisal

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK Sarah Davis, Senior Lecturer in Health Economics, ScHARR, University of Sheffield, Sheffield, UK John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK Alison Scope, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK Kate Ennis, Research Associate, ScHARR, University of Sheffield, Sheffield, UK Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK
Correspondence Author	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
Date completed	12/03/2019

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 18/54/13.

Declared competing interests of the authors

None of the authors have any conflicts of interest to declare.

Acknowledgements

We would like to thank Paul Tappenden, ScHARR, for providing comments on the draft report and Andrea Shippam, Programme Manager, ScHARR, for providing administrative support and in preparing and formatting the report.

Rider on responsibility for report

This report was commissioned by the NIHR HTA Programme. The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Martyn-St James M, Davis S, Stevens J, Scope S, Ennis K, Wong J, Alhasso A, Birch B, Chitale S, Lewi H. Pentosan polysulfate sodium for treating bladder pain syndrome: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2019.

Contributions of authors

Marrissa Martyn-St James and Alison Scope summarised and critiqued the clinical effectiveness data reported within the company's submission. Sarah Davis and Kate Ennis critiqued the health economic analysis submitted by the company. John Stevens critiqued the statistical analyses undertaken by the company. Ruth Wong critiqued the company's search strategy. Ammar Alhasso, Brian Birch, Sudhanshu Chitale, and Henry Lewi, provided clinical advice to the ERG throughout the project. All authors were involved in drafting and commenting on the final report.

CONTENTS

List of tables

List of figures

Abbreviations

AE	Adverse event
AiC	Academic in confidence
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BIs	Bladder instillations
BPS	Bladder pain syndrome
BSC	Best supportive care
CI	Confidence interval
CS	Company submission
DES	Discrete event simulation
DMSO	Dimethyl sulphoxide
DSA	Deterministic sensitivity analysis
DSU	Decision Support Unit
ED	Emergency department
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EQ-5D	EuroQol 5 Dimensions
EQ-5D-3L	EuroQol 5 Dimensions 3-level version
EQ-5D-5L	EuroQol 5 Dimensions 5-level version
ERG	Evidence Review Group
FDA	Food and Drug Administration
GAG	Glycosaminoglycan
GP	General practice
GRA	Global response assessment
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
IC/BPS	Interstitial cystitis/bladder pain syndrome (patients with bladder pain syndrome with Hunner's lesions and/or glomerulations)
ICER	Incremental cost effectiveness ratio
ICPI	Interstitial Cystitis Problem Index
ICSI	Interstitial Cystitis Symptom Index
ITC	Indirect treatment comparison
ITU	Intensive therapy unit
MIMS	Monthly Index of Medical Specialities
mL	Millilitre

mm	Millimetre
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases
ONS	Office for National Statistics
PSA	Probabilistic sensitivity analysis
PPS	Pentosan polysulfate sodium
PSS	personal and social services
QALY	Quality-adjusted life-year
RCT	Randomised controlled trial
RR	Relative risk
SD	Standard deviation
SLR	Systematic literature review
STA	Single Technology Appraisal
VAS	Visual analogue scale
VBA	Visual Basic for Applications

1 SUMMARY

1.1 Critique of the decision problem in the company's submission

The Evidence Review Group (ERG) considers the company's description of the underlying health problem in the company's submission (CS) to be appropriate and relevant to the decision problem set out in the final scope issued by the National Institute for Health and Care Excellence (NICE). The decision problem assesses pentosan polysulfate sodium (PPS) (Elmiron®) for treating interstitial cystitis/bladder pain syndrome (patients with bladder pain syndrome with Hunner's lesions and/or glomerulations) (IC/BPS). In accordance with the NICE scope, the target population in the CS is people with IC/BPS. The comparator in the NICE scope is bladder instillations (BIs). For people in whom this treatment is inappropriate, unsuccessful, or cannot be tolerated, established clinical management without PPS (also referred to as best support care [BSC]) is the comparator. However, the CS only includes clinical effectiveness evidence for bladder instillations containing sodium hyaluronate, sodium chondroitin sulphate, or a combination of both. Clinical advice received by the ERG indicates that there is some variability in the availability of, and ingredients used, in locally prepared instillations across hospitals, but that these instillations could be appropriate and relevant comparators; however, it is unclear how frequently they are used. The company's clarification response stated that locally prepared instillations are not included because of their relatively infrequent use in the UK, the heterogeneity of the different 'cocktails', and the difficulty in sourcing relevant evidence for their use in IC/BPS.

1.2 Summary of clinical effectiveness evidence submitted by the company

The key clinical effectiveness evidence in the CS for PPS in IC/BPS was based primarily on four randomised controlled trials (RCTs). The trial populations in the four RCTs relate to patients who have IC/BPS. All four RCTs compared PPS to placebo (PBO). Two RCTs comparing sodium chondroitin sulphate instillations (Uracyst®) to PBO in BPS were also included which were used to construct an indirect treatment comparison (ITC) based on the Bucher method between PPS and sodium chondroitin sulphate instillations for use in the economic model.

The four RCTs of PPS in IC/BPS were relevant to the decision problem outlined in the final NICE scope.

Two of the RCTs of PPS in IC/BPS reported that the between-group difference in the proportions of patients with a >50% improvement in global response assessment (GRA) at three months was statistically significant in favour of PPS. However, in one RCT the between-group difference in the proportions of patients with a GRA score of six to seven at three months was reported as not statistically significant. As GRA was not assessed in one RCT, the company used non-VAS pain data

at three months from the RCT as a proxy for GRA in their meta-analysis for this outcome. The between-group difference in the proportions of patients with a >50% improvement in non-VAS pain in this RCT was reported as statistically significant. The between-group difference in the proportions of patients with a >50% improvement in non-VAS pain at three months was also reported as statistically significant in one other RCT, but the between-group difference in mean non-VAS pain scores was reported as not statistically significant in two RCTs.

In the company's pairwise meta-analysis of PPS in IC/BPS, the pooled relative risk (RR) for GRA at three months across the four RCTs of PPS in IC/BPS was 2.09 (95% CI: 1.47 to 2.97, fixed effect). These results were used in the economic model. In the company's pairwise meta-analysis of Uracyst® in BPS, the pooled RR for GRA at trial follow-up across the two Uracyst® RCTs was 1.39 (95% CI: 0.88 to 21.7, fixed effect). These results were also used in the economic model. The between-group difference in the proportions of patients with a GRA score of six to seven at the trial follow-up was reported as not statistically significant by both of the Uracyst® RCTs.

In PPS in IC/BPS, the between-group difference in the O'Leary-Sant Interstitial Cystitis Symptom Index and Problem Index scores at three months were both reported as not statistically significant by one RCT.

Across the RCTs of PPS in IC/BPS, no statistically significant between-group differences were reported at three months in mean: daily urinary frequency (two RCTs), urinary volume and void outcomes (three RCTs), and nocturia (two RCTs). One RCT did not report whether the between-group difference at three months was significant or not for mean urinary volume and void outcomes, or mean nocturia.

Safety data for PPS were presented from each of the individual RCTs of PPS in IC/BPS, and the company concluded that PPS is well tolerated. Common adverse events in the SmPC are: headache, dizziness, nausea, diarrhoea, dyspepsia, abdominal pain, abdominal enlargement, rectal haemorrhage, peripheral oedema, alopecia, back pain, asthenia, and pelvic pain. However, clinical advice received by the ERG based on named patient use is that AEs are rare with PPS.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The ERG considers the searches for clinical effectiveness evidence reported in the CS to be adequate, and believes the included RCTs of PPS to be relevant to the NICE decision problem. The ERG notes that there have been no other published independent studies validating the results of these RCTs since the four pivotal RCTs of PPS in IC/BPS were conducted.

The eligibility criteria applied in the selection of evidence for clinical effectiveness were considered by the ERG to be reasonable and consistent with the decision problem outlined in the final NICE scope. However, the company chose to also include RCTs of PPS and comparators that were in the broader BPS population (patients with bladder pain syndrome but without Hunner's lesions and/or glomerulations). Primary endpoints and selected analyses for clinical efficacy were appropriate.

The quality of the included RCTs was assessed using well-established and recognised criteria by NICE. However, the company only quality assessed the RCTs of PPS. Quality assessment of comparator treatment RCTs was not undertaken by the company and was therefore undertaken by the ERG.

The ERG notes limitations in the reporting of outcome data in the PPS RCTs trial reports. Interval estimates (CIs) were not reported and, where between-group differences were reported as not statistically significant, *p*-values were often not reported.

Three of the RCTs of PPS in IC/BPS were considered by the ERG to be of good methodological quality. However, the ERG considers one RCT to be unclear regarding: allocation concealment, details of who was blinded, and the number of patients withdrawing from treatment groups.

The ERG considers that the company's overview of the safety evidence from the RCTs of PPS in IC/BPS reported in the CS and the company's conclusion that PPS is well tolerated to be reasonable.

The ERG has concerns with the two Uracyst® RCTs used for the ITC with PPS. Both of the Uracyst® RCTs were in the broader BPS population (without Hunner's lesions and/or glomerulations), and both compared Uracyst® to a placebo bladder instillation (not a tablet).

The ERG has some concerns with the meta-analyses that were performed by the company and reported in the CS (analysis using risk difference, assessment of heterogeneity, application of a fixed effect model). However, the ERG accepts the company's argument that an unbiased comparison between PPS capsules and all relevant comparators was not possible using a conventional network meta-analysis.

In order to include comparisons of PPS against all comparators listed in the NICE scope, the company provided an ITC between PPS and Uracyst® linked by the placebos. The ERG has some concerns with the method for the ITC (based on the Bucher method) and would prefer a simultaneous comparison between treatments using a Bayesian network meta-analysis as: (i) the Bucher approach allows for separate and unrelated meta-analyses for the effect of PPS versus placebo and the effect of

Uracyst® versus placebo whereas a single model incorporates a common random effect; (ii) the posterior distribution for the effect of PPS versus Uracyst® will not follow any standard parametric distribution whereas the Bucher approach involves an assumption of asymptotic normality when making inferences, and; (iii) the relative treatment effects of PPS versus placebo and Uracyst® versus placebo will be correlated and this will induce correlation between absolute responses to treatment when combined with an external estimate of the baseline response.

1.4 Summary of cost effectiveness submitted evidence by the company

The CS includes a *de novo* economic analysis, which compares PPS to BIs in patients able to receive BIs and PPS to BSC in patients unable to receive BIs. In both cases, the population matches that specified in the marketing authorisation and the NICE scope. The model uses patient-level simulation to estimate expected costs and quality-adjusted life years (QALYs) over a 20-year time-horizon using a discount rate of 3.5% per annum. The company's economic analysis adopts an NHS perspective for costs and benefits are restricted to patients. Benefits to carers and costs falling on personal and social services (PSS) were not considered relevant.

The company submitted a revised model following the clarification request and it is this model that is referred to throughout the report unless otherwise specified. The revisions were mainly corrections of errors in the implementation of the model.

The company's model uses a discrete event simulation (DES) framework, with the main events being a response check at 6 months, a discontinuation event which applies only to responders and death from all-cause mortality. Patients who have responded at 6 months are assumed to remain on their first-line treatment until discontinuation or death. Patients who do not respond are assumed to switch to second-line treatment; this is assumed to be BIs for those patients who are able to receive BIs, and BSC for those unable to receive BIs. Patients having BIs as first-line therapy also have events for each individual BI administration, allowing the frequency of the BIs to vary over time. BIs given as second-line therapy are modelled based on the mean number of administrations per annum without modelling each administration as a separate event.

The key model inputs are the response rates for each first-line treatment option, costs and utilities for responders and non-responders and time to treatment discontinuation for first-line treatment. The response rates were based on the company's systematic review and meta-analyses. The comparison between PPS and BIs was based on a simple unadjusted indirect comparison using the Bucher method. The costs and utilities for responders and non-responders are estimated based on the expected Interstitial Cystitis Symptom Index (ICSI) scores for responders and non-responders, estimated using data from the PPS arm of one RCT. The relationship between ICSI score and costs and utilities has

been estimated from regressions fitted to data from a patient survey. Utilities were estimated by mapping from the EQ-5D-5L responses obtained in the patient survey to the EQ-5D-3L UK valuation set. Disease costs were estimated by combining resource use data obtained in the patient survey with NHS reference costs. In the regression applied in the model, disease costs are dependent only on age and ICSI score, but utilities are also dependent on whether patients have received BIs in the past 6 months. Time to treatment discontinuation for PPS has been estimated from a published observational study with long-term discontinuation rates extrapolated based on a parametric survival analysis. The time to treatment discontinuation for BIs has been assumed to be equivalent to that for PPS. Life expectancy in the model was based on general population mortality rates for all treatment options with none of the treatments having any impact on mortality. In addition to disease-related costs that depend on the expected ICSI score, treatment-related costs include acquisition costs for PPS and BIs and administration costs for BIs. Costs and health impacts related to AEs were not included in the model.

In the population able to receive BIs, the company's revised deterministic model estimated that PPS would generate 0.25 additional QALYS in comparison to BIs, at an additional cost of [redacted]; giving an ICER of [redacted] per QALY gained. The base-case probabilistic ICER for PPS versus BIs was [redacted] per QALY gained with a 0.54 probability of PPS being cost-effective compared to BI at a willingness-to-pay threshold of £20,000 and a 0.61 probability of PPS being cost-effective compared to BI at a willingness-to-pay threshold of £30,000.

In the population unable to receive BIs, the company's revised deterministic model estimates that PPS generates 0.32 additional QALYS in comparison to BSC, at an additional cost of [redacted]; giving an ICER of [redacted] per QALY gained. The company's base-case probabilistic ICER for PPS versus BSC was [redacted] per QALY gained, with a 0.15 probability of being cost-effectiveness at the £20,000 willingness to pay threshold and a 0.33 probability of being cost-effective at the £30,000 willingness to pay threshold.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The company's model is generally in line with the NICE reference case, with the main significant deviations being: 1) that the comparison between PPS and BIs is based on a simple unadjusted indirect comparison using the Bucher method; and 2) that the estimates of clinical effectiveness for BIs versus placebo were taken from the broader population with BPS rather than the population with IC/BPS that matches the licensed indication for PPS. In addition, the ERG believes that a lifetime horizon would have been preferable to the company's 20-year time horizon.

The key areas of concern identified by the ERG were:

- The application of a utility decrement for patients receiving BIs estimated from the patient survey which the ERG did not consider robust given that the handling of missing data on BI usage had not been adequately explored in the analysis of the survey data.
- Uncertainty surrounding the likely rate of response in patients receiving BSC in clinical practice which affects the absolute difference in response attributable to PPS in the model.
- Inconsistent assumptions around the durability of response in those receiving BSC and those receiving either PPS or BIs.
- The assumption that 4-weekly administration of BIs (i.e. 13 per annum) continues indefinitely when the ERG believes that the frequency of administration is likely to fall over time as the spacing between doses is increased to the longest interval that patients can tolerate.
- Underestimation of discontinuation rates from Hanno *et al.* (1997) which affects the lifetime treatment costs, particularly for the comparison of PPS versus BSC.
- The assumption that patients who do not respond to BSC have some long-term persistent utility gain and cost savings relative to baseline.
- The assumption that the long-term cumulative rate of response to second-line BIs is equivalent to the short-term response to first-line BIs.
- Low rates of self-administration for BIs which may overestimate costs relative to established clinical practice in some parts of the NHS.
- The simplistic approach to estimating expected ICSI scores for responders and non-responders.

1.6 ERG commentary on the robustness of evidence submitted by the company

1.6.1 Strengths

The ERG considers the data on clinical effectiveness in the CS to be reasonably well-reported and that three of the four pivotal RCTs of PPS in IC/BPS are of reasonably good quality. However, there are aspects of uncertainty surrounding one RCT of PPS in IC/BPS.

The safety profile submitted by the company is based on the adverse events reported in the four RCTs of PPS in IC/BPS. Common adverse events (AEs) listed in the SmPC are: headache, dizziness, nausea, diarrhoea, dyspepsia, abdominal pain, abdominal enlargement, rectal haemorrhage, peripheral oedema, alopecia, back pain, asthenia, and pelvic pain. However, clinical advice based on named patient use received by the ERG is that AEs with PPS are uncommon.

The company provides a valid argument that an unbiased comparison between PPS capsules and all relevant comparators was not possible using a conventional network meta-analysis. Nevertheless, as required for the economic evaluation, the company provided an unadjusted ITC between PPS and Uracyst® linked by the placebos. In the absence of any direct measure of health-related quality of life

from the RCTs, the company has conducted a patient survey to estimate utility values derived from the EQ-5D that comply with the NICE reference case.

1.6.2 Weaknesses and areas of uncertainty

The four pivotal RCTs of PPS in IC/BPS were conducted between 1987 and 2003, and there is commonality across trial investigators. The FDA queried the independence of investigators across two of the RCTs, along with the possibility of a treatment-by-investigator effect for one of seven study centres in one RCT. To date, there has been no further, independent, published study validating the results of the four RCTs of PPS in IC/BPS.

The ERG has concerns with the pairwise meta-analyses that were performed by the company and reported in the CS (analysis using risk difference, assessment of heterogeneity, application of a fixed effect model). There are also concerns with the method for the ITC (based on the Bucher method) and the ERG would prefer a simultaneous comparison between treatments using a Bayesian network meta-analysis.

The likely rate of response in patients receiving BSC without either PPS or BIs in clinical practice is uncertain and the estimates of cost-effectiveness are very sensitive to this rate. It is unclear what costs and utilities values should be assumed in the model for patients who respond to BSC. The relationship between prior use of BIs and utility is not considered to be robust given that the handling of missing data on BI usage had not been adequately explored in the analysis of the survey data. The CS does not contain any data describing the frequency of BIs in clinical practice and whether this decreases over time, or any data on the rate of self-administration with BIs. Several strong assumptions have had to be made in the company's model to deal with a lack of data on: (a) long term discontinuation rates for BIs; (b) the relative effectiveness of BIs and PPS; (c) the effectiveness of BIs in the population with IC/BPS; (d) the long-term response rate for patients cycling through multiple BIs after failing to respond to a first-line BI treatment, and (e) the relationship between ICSI scores and response to treatment.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG undertook seven sets of exploratory analysis by implementing changes to the company's revised model. The ERG's preferred base-case incorporates all of these seven model amendments:

1. Use of all discontinuations reported by Hanno *et al.* (1997) for the time to treatment discontinuation survival analysis.
2. Switch to 6-weekly dosing for first-line BIs after the first year of treatment and 6-weekly for all second-line BIs (affects PPS vs BI only).
3. Use of regression for utility based on ICSI scores which excludes term for prior usage of BI.

4. Use of a lifetime horizon.
5. Return to baseline utilities for non-responders when BSC is second-line option (affects PPS versus BSC scenario only).
6. Remove assumption that response stops at 12 month for responders to BSC (PPS versus BSC scenario only as already implemented in PPS versus BI base-case).
7. Use of log-normal distribution to model the time to treatment discontinuation

The exploratory analysis demonstrated that the ICER for PPS vs BIs was most sensitive to changes in the frequency of BIs instillations (ICER increased to [redacted] per QALY gained) and the use of the utility regression that excludes the coefficient for recent BI usage (ICER increased to [redacted] per QALY gained). The ICER for the ERG's preferred base-case was [redacted] per QALY gained.

The exploratory analysis demonstrated that the ICER for PPS vs BSC was most sensitive to the removal of the assumption that the BSC response recedes at 12 months (ICER increased to [redacted] per QALY gained) and changes to the data on time to treatment discontinuation (ICER increased to [redacted] per QALY gained when the exponential distribution was used and [redacted] per QALY gained when the log-normal distribution was used). The ICER reduced significantly to [redacted] per QALY gained when assuming that non-responders on BSC return to base-line values for utility and costs. Overall, the ICER for the ERG's preferred base-case was [redacted] per QALY gained.

The ERG also conducted further sensitivity analyses around their preferred base-case to explore the impact of several data inputs and assumptions that remain uncertain. This produced ICERs ranging from [redacted] [redacted] per QALY gained for PPS vs BIs and ICERs ranging from [redacted] [redacted] per QALY gained for PPS vs BSC. The ICERs were particularly sensitive to uncertainty regarding the proportion of patients who would be expected to respond to BSC and uncertainty regarding the likely rate of self-administration of BIs in clinical practice.

2 BACKGROUND

This report provides a review of the evidence submitted by Consilient Health in support of pentosan polysulfate sodium (PPS) (Elmiron®) for treating interstitial cystitis/bladder pain syndrome (IC/BPS). It considers both the original company submission[1] (CS) received on 9th January 2019 and a subsequent response to clarification questions supplied by Consilient Health on 13th February 2019.[1]

2.1 Critique of company's description of underlying health problem

The Evidence Review Group (ERG) considers the company's description of the underlying health problem in the company's submission[1] (CS) to be appropriate, mostly up-to-date and relevant to the decision problem set out in the final National Institute for Health and Care Excellence (NICE) scope.[2] The ERG provides a brief summary of the underlying health problem in this section.

Clinical features and nomenclature

The European Association of Urology 2018 guidelines on chronic pelvic pain describes bladder pain syndrome (BPS) as a chronic bladder condition characterised by persistent or recurrent pain, accompanied by at least one other symptom, such as pain worsening with bladder filling and day-time and/or night-time urinary frequency.[3] Other terms that have been used, but that are no longer recommended by the European Association of Urology include: interstitial cystitis (IC), painful bladder syndrome (PBS), and PBS/IC or BPS/IC.[3] The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) criteria for the diagnosis of interstitial cystitis/bladder pain syndrome (IC/BPS) includes a diagnosis of IC based on glomerulations (haemorrhages in the bladder wall) or Hunner's lesions (distinctive inflammatory lesions that rupture the bladder lining) on cystoscopic examination.[4] The CS[1] uses the term BPS to describe patients meeting the broader symptomatic criteria of chronic bladder pain,[5] and IC/BPS to describe those with symptoms of BPS who also have glomerulations and/or Hunner's lesions and who comprise the indicated population for PPS.

Aetiology

The aetiologies of both BPS and IC/BPS are unknown, although several theories have postulated, including that of a deficient glycosaminoglycan (GAG) layer in the bladder.[6]

Prevalence

In the UK, BPS may affect approximately 400,000 people, 90% of whom are women.[7] and is more common in women than men. Up to 50% of patients with symptoms of BPS will have spontaneous resolution in time.[8] In Europe, estimates of the prevalence of BPS associated with inflammation in the bladder (for example, characterised by Hunner's lesions or glomerulations) range from 0.3 to 10.2 per 10,000 patients.[9-11]

Diagnosis

Clinical diagnosis of BPS is often made once specific causes such as infection and malignancy have been ruled out.[12, 13] Diagnosis is made using symptoms, examination, urine analysis and urine culture (to rule out a urinary tract infection), cystoscopy with or without hydrodistension (to rule out bladder cancer, vesical stones, urethral diverticula and intravesical foreign bodies), and biopsy (to exclude other pathologies).[8]

2.2 Critique of company's overview of current service provision

The ERG considers the company's overview of current service provision to be reasonable, in that the company acknowledges that there is currently no NICE guidance on the management of BPS or IC/BPS. The company presents a proposed patient/treatment pathway for IC/BPS. The ERG provides a brief summary of this in this section.

Proposed patient/treatment pathway

For people with IC/BPS, an advisory board to the company concluded that PPS or bladder instillations are second-line treatments after standard management (e.g. analgesics, hydroxyzine, lifestyle/dietary advice, bladder retraining).[1] An advisory board to the company proposed both a patient and a treatment pathway. The proposed patient pathway by reproduced from the CS,[1] is presented in Figure 1. The proposed treatment pathway reproduced from the CS,[1] is presented in Figure 2.

The company's advisory board also concluded that bladder instillations include commercially available instillations, such as sodium hyaluronate (Cystistat® , Hyacyst®) and sodium chondroitin sulphate (Uracyst® , Gepan®), or locally prepared instillations using ingredients (off-label) such as heparin, lignocaine, sodium bicarbonate or hydrocortisone (CS, page 23). However, whilst the proposed treatment pathway presented in the CS included locally prepared instillations, evidence for these was not included in the CS.

During the clarification process, the ERG asked the company why these treatments were not included in the CS. In response, the company stated that these locally prepared instillations, also known as 'bladder cocktails', can vary by site and include commonly used drugs indicated for other conditions. Further, that these have not been included in the company's submission because of their relatively infrequent use in the UK, the heterogeneity of the different cocktails, and the difficulty in sourcing relevant evidence of their use in IC/BPS.[1]

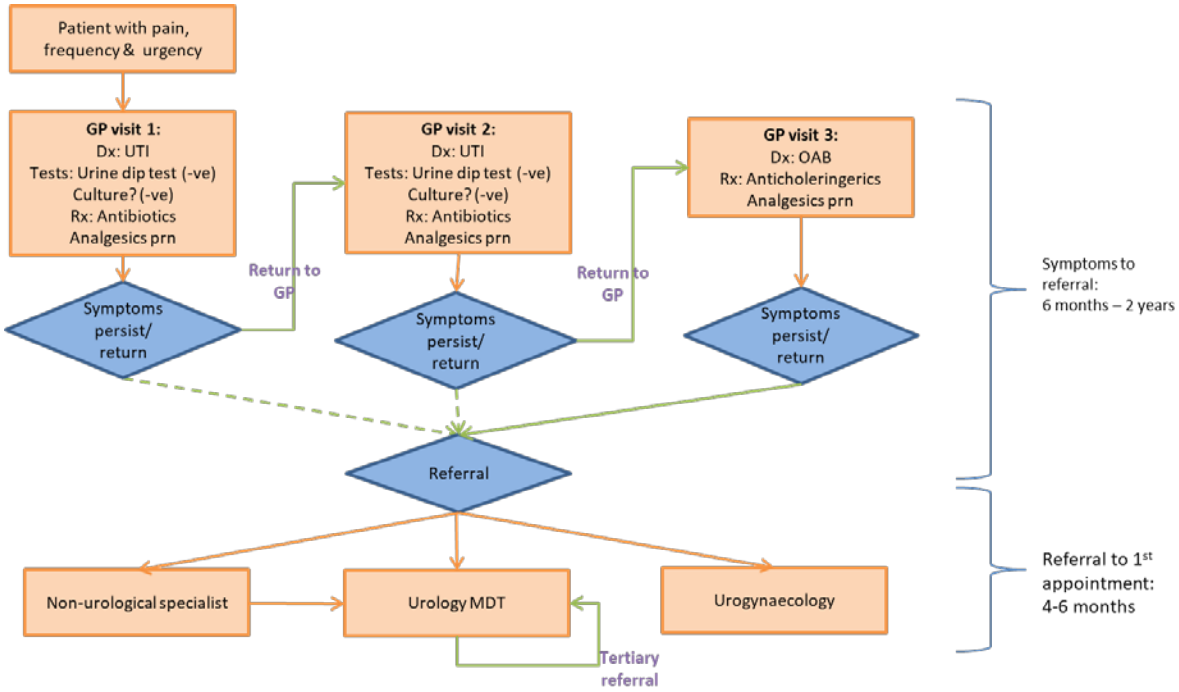
The advisory board to the company also concluded that sodium hyaluronate/sodium chondroitin sulphate (iAluRil®) is often not used until later in the pathway as a third-line treatment if other

instillations are unsuccessful (CS, page 23). The company's advisory board also noted that prior to the UK launch of licensed Elmiron® (PPS) in September 2018, oral PPS was only available as an unlicensed special import (CS, page 24). The advisory board to the company concluded that surgery including urinary diversion, bladder reconstruction (i.e., augmentation), and cystectomy, is considered as a last resort (CS, page 24) and that the proportion of IC/BPS patients receiving surgery is low (2%) (CS,[1] Figure 2).

Clinical advice received by the ERG on the proposed patient/treatment pathway varied. Some clinical experts expressed a wish to use PPS before BIs as it is less invasive, whilst others felt that it would be used after failure of BIs. Clinical advice received by the ERG on the experience of using PPS and its availability off-label varied. There was no consensus on the use of locally prepared bladder instillations containing heparin, lignocaine, sodium bicarbonate or hydrocortisone; or the use of botulinum toxin A in treating IC/BPS. However, there was consensus that the proportion of IC/BPS patients receiving surgery in the UK is very low (2% to 5%).

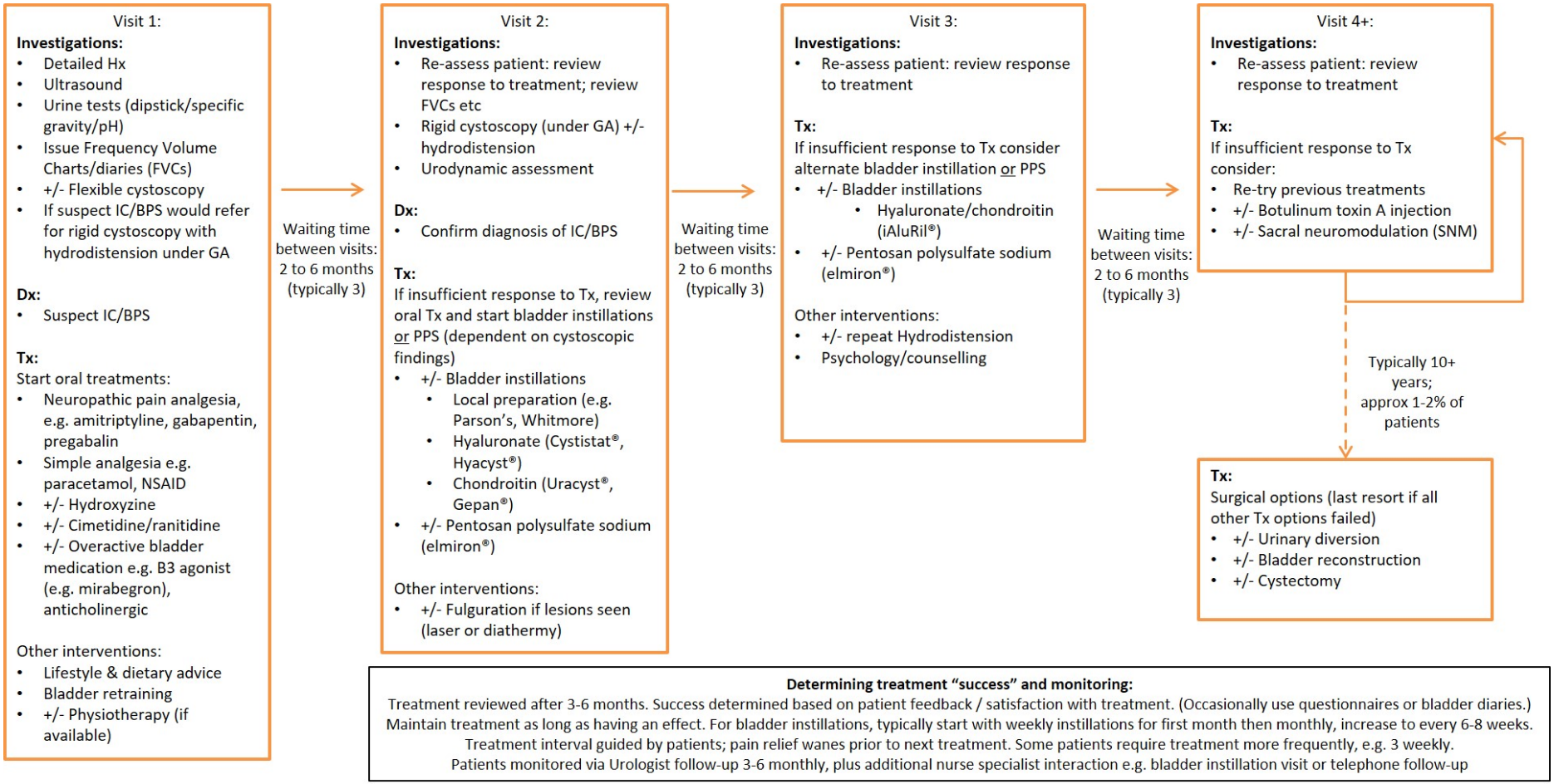
The advisory board to the company suggested that the number of BPS patients for whom BIs are contraindicated or who refuse bladder instillations is <5%.[1] The ERG's clinical advisors believe this to be reasonable.

Figure 2 of the CS (Figure 2) states that bladder instillations typically start weekly for the first month, then monthly, then decrease in frequency to every six to eight weeks. Clinical advice received by the ERG on the proposed frequency of instillations varied, but was generally consistent with weekly instillations for the first four to six weeks, prior to lengthening the treatment interval.



Reproduced from the CS page 25.[1]

Figure 1: Patient pathway for IC/BPS proposed by the advisory board to the company presented in the CS (Figure 1)



Reproduced from the CS page 26.[1]

Figure 2: Treatment pathway for IC/BPS proposed by the advisory board to the company presented in the CS (Figure 2)

3 CRITIQUE OF COMPANY'S DEFINITION OF THE DECISION PROBLEM

3.1 Population

Pentosan polysulfate sodium (PPS) (Elmiron® , Consilient Health) has a marketing authorisation in the Europe for treating IC/BPS. The target population in the company's decision problem matches the population described in the final NICE scope which is 'adults with bladder pain syndrome characterised by either glomerulations or Hunner's lesions with moderate to severe pain, urgency, and frequency of micturition',[2] which is narrower than the marketing authorisation.

The key clinical evidence submitted by the company is derived from four randomised controlled trials (RCTs) of PPS in IC/BPS.[14-17] These RCTs all recruited patients with glomerulations and/or Hunner's lesions and were undertaken in the United States. Clinical advice received by the ERG suggested that the populations in these RCTs are generally comparable to the UK IC/BPS population. The company also included two additional RCTs of PPS in the broader BPS population that did not include a cystoscopic evaluation for glomerulations or Hunner's lesions at baseline.[18, 19] These two RCTs did not contribute to the pairwise meta-analysis of global response used in the company's base-case economic model, but did contribute to other meta-analyses in the clinical section of the CS. In addition, the impact on the cost-effectiveness estimates of including them in the meta-analysis used to estimate the rate of response for PPS in the company's model was examined in a scenario analysis. These two RCTs are not considered further in this section of the ERG report, but are summarised briefly in Section 4.2.5.

3.2 Intervention

The intervention evaluated in the CS is Elmiron® (pentosan polysulfate sodium, PPS), a semi-synthetic heparin-like substance that resembles glycosaminoglycans (GAGs). Although its exact mechanism of action is unclear, PPS is hypothesised to bind to the damaged GAG layer in the bladder, which protects the bladder by reducing the adherence of bacteria to the mucosal lining, in turn reducing inflammation. In addition to its anti-inflammatory activity, PPS may also have a barrier function instead of the damaged urothelial mucus.[1] The intervention matches that in the NICE scope.[2]

PPS received a positive opinion from the Committee for Medicinal Products for Human Use (CHMP) on the 23rd March 2017 for the treatment of IC/BPS, and received EMA marketing authorisation June 2017.[8]

The Summary of Product Characteristics (SmPC)[8] reports that PPS is contraindicated in patients who actively bleed (excluding menstruation).[8]

The SmPC recommends that patients undergoing invasive procedures or having signs/symptoms of underlying coagulopathy or other increased risk of bleeding should be evaluated for haemorrhagic events, and patients who have a history of heparin or PPS induced thrombocytopenia should be carefully monitored.[8] The ERG's clinical advisors agreed with this.

Common adverse events (AEs) listed in the SmPC are: headache, dizziness, nausea, diarrhoea, dyspepsia, abdominal pain, abdominal enlargement, rectal haemorrhage, peripheral oedema, alopecia, back pain, asthenia, and pelvic pain.[8] Clinical advice received by the ERG from experience of using PPS on a named patient basis is that AEs are rare.

PPS is administered orally three times per day. The list price for PPS is £450.00 per pack (90 x 100 mg capsules). The cost-effectiveness results presented by the company are based on the list price.

3.3 Comparators

Two comparators are listed in the final NICE scope: (i) bladder instillations, and (ii) for people for whom bladder instillations are inappropriate, cannot be tolerated or are unsuccessful: established clinical management without PPS or bladder instillations (including medicines that do not currently have a marketing authorisation in the UK for this indication).[2]

Whilst the CS reports that bladder instillations include commercially available instillations, such as sodium hyaluronate (Cystistat® , Hyacyst®) and sodium chondroitin sulphate (Uracyst® , Gepan®), or locally prepared instillations using ingredients (off-label) such as heparin, lignocaine, sodium bicarbonate or hydrocortisone, only evidence relating to sodium hyaluronate, sodium chondroitin sulphate or a combination of the two (iAluRil®) was searched for.[1] Clinical advice received by the ERG on the use of off-label instillations in IC/BPS varied. There was no consensus regarding the use of locally prepared instillations using ingredients (off-label) such as heparin, lignocaine, sodium bicarbonate or hydrocortisone. The ERG sought clarification (question A4) with the company regarding these treatments not being including in the CS. The company's clarification response stated these locally prepared instillations, also known as 'bladder cocktails', can vary by site and include commonly used drugs indicated for other conditions. Further, that these have not been included in the company's submission because of their relatively infrequent use in the UK, the heterogeneity of the different cocktails, and the difficulty in sourcing relevant evidence for their use in IC/BPS.[1]

Two RCTs in sodium chondroitin sulphate (Uracyst®) in patients with BPS[20, 21] were included in the ITC presented in the CS.[1]

3.4 Outcomes

The outcomes in the decision problem in the CS are:

- bladder pain, response to treatment (e.g., Global Response Assessment [GRA], a standardised outcome in IC/BPS),
- severity of symptoms,
- urinary urgency,
- urinary frequency,
- nocturia,
- adverse effects of treatment, and
- health-related quality (HRQoL) of life.[1]

These outcomes match those in the NICE scope.[2]

Across the four included RCTs of PPS in IC/BPS,[14-17] the CS[1] presents outcome data on: GRA; the Interstitial Cystitis Symptom Index (ICSI);[22] Interstitial Cystitis Problem Index (ICPI), non-VAS pain outcomes (not defined in the CS), urinary frequency, void/volume outcomes, nocturia, and adverse events.

Additional outcomes of maximum bladder capacity, cystoscopic outcomes, cystometric outcomes, and mast cell count are also reported for one RCT in the broader BPS population.[18] The CS notes that the advisory board recommended that the comparability of bladder capacity at baseline across trials be assessed. However, this was not included in the CS. The company's clarification response included baseline bladder capacity reported across the included RCTs.[1]

The CS states that the measures of GRA from the four RCTs of PPS in IC/BPS[14-17] are equivalent.[1] Clinical advice received by the ERG was generally in agreement with this. Clinical advice received by the ERG also indicated the possibility of a 20% to 40% response to BSC in clinical practice for this outcome in clinical practice.

3.5 Other relevant factors

Equity

The CS reports that the evaluation does not include weighting of quality-adjusted life years (QALYs) (CS,[1] Table 39).

Adherence

Adherence to treatment is not measured in the CS.[1] The CS describes the hypothesised mechanism of action for PPS in binding to the GAG layer of the bladder, thus reducing adherence of bacteria and reducing inflammation (CS, page 14). The ERG's clinical advisors suggest that PPS may take up to three months to be effective. The clinical advisors stated that IC/BPS patients are advised to continue with other current treatment which will continue to have some therapeutic effect after starting PPS and that IC/PPS patients tend to stay on a treatment that is working and that stopping treatment may result in an IC/BPS symptoms flare.

Ongoing studies

The company searched appropriate sources to identify ongoing studies; the CS states that no ongoing studies of PPS in IC/BPS were identified (CS, Section B.2.11).[1]

Patient Access Scheme

The CS reports that a Patient Access Scheme for PPS is not applicable (CS, Table 2).[1]

4 CLINICAL EFFECTIVENESS

This section presents a review of the clinical evidence reported in the CS[1] for pentosan polysulfate sodium (PPS) for treating interstitial cystitis/bladder pain syndrome (patients with bladder pain syndrome with Hunner's lesions and/or glomerulations) (IC/BPS). The RCTs are presented in evidence tables in the CS and in this ERG report in reverse chronological order (most recent first).

4.1 Critique of the methods of review(s)

The clinical evidence provided in the CS comprises a systematic review of RCTs of PPS for both IC/BPS (four RCTs[14-17]) and BPS (two RCTs, [18, 19] summarised in Section 4.2.5), a pairwise meta-analysis of four RCTs of PPS in IC/BPS[14-17], a pairwise meta-analysis of two RCTs in sodium chondroitin sulphate instillations (Uracyst®) in BPS,[20, 21] and an ITC of PPS in IC/BPS compared to Uracyst® in BPS. Safety evidence provided in the CS comprises a narrative synthesis of four RCTs of PPS in IC/BPS [14-17] and two RCTs of Uracyst® in BPS.[20, 21]

4.1.1 Searches

The company performed a systematic literature review (SLR) to identify all clinical and safety studies of pentosan polysulfate sodium and its comparators for the treatment of patients with or without cystitis or bladder pain.

For the original searches, several electronic bibliographic databases were searched in June 2018 including MEDLINE [via Ovid], MEDLINE Epub Ahead of Print, in Process [via Ovid], Embase [via Ovid], Cochrane Database of Systematic Reviews [via Wiley], Cochrane Central Register of Controlled Trials and the Health Technology Assessment database [via Wiley], Database of Abstracts of Reviews of Effects [via Wiley] and the Health Technology Assessment Database [via Wiley]. The company did not search conference proceedings websites or databases (clarification question A3) for unpublished studies. However, the company searched two key clinical trials registers (clinicaltrials.gov, WHO International Clinical Trials Registry Platform).

In Appendix D (RCTs and non-RCTs), the company only reported the full literature search strategies for identifying RCTs. The company's response to clarification question A3 stated that the comparators heparin, lignocaine, sodium bicarbonate and hydrocortisone were excluded from the clinical effectiveness search because of infrequent use in the UK, the heterogeneity of the mixtures and usage and the difficulty in sourcing relevant data

In response to clarification question A2, the company provided search strategies for the clinical effectiveness evidence search for non-randomised studies (reported in Table 27 of the CS). It is

unclear why the company only searched one electronic database (PubMed via NIH) and two other web sources DIMDI and MedPilot rather than Embase and Cochrane Library. The company performed a high precision search of interstitial cystitis combined with pentosan sulphuric polyester in PubMed but did not report on the strategy for searching DIMDI and MedPilot. The ERG was unable to assess the adequacy of the non-RCT searches. For the reasons described above, the ERG was also unable to assess the adequacy of the searches for Medline and Cochrane Library.

4.1.2 Inclusion criteria

The inclusion and exclusion criteria for the systematic review are reported in the CS[1] are in accordance with the NICE scope,[2] with the exception of locally prepared instillations using ingredients (off-label) such as heparin, lignocaine, sodium bicarbonate or hydrocortisone. The ERG sought clarification from the company regarding the exclusion of these treatments from the CS.

A copy of the inclusion and exclusion criteria, reproduced from the CS[1] are presented in Table 1.

Table 1: Inclusion and exclusion criteria in systematic review search strategy (reproduced from Table 65 of the CS)

Characteristics	Inclusion criteria	Exclusion criteria
Population	Adult patients (≥18 years) with interstitial cystitis/bladder pain syndrome (IC/BPS) or BPS	Paediatric patients (<18 years)
Interventions	Elmiron® (pentosan polysulfate sodium/sodium pentosan polysulfate)	NA
Comparators	• Cystistat® (sodium hyaluronate/hyaluronic acid 0.08%) • Hyacyst® (sodium hyaluronate/hyaluronic acid 0.08% or 0.24%) • Gepan® (sodium chondroitin sulphate 0.2%) • Uracyst® (sodium chondroitin sulphate 2%) • iAlurRl® (hyaluronic acid/sodium hyaluronate 1.6% and sodium chondroitin sulphate 2%) • Placebo	Studies not comparing the intervention with a comparator or studies not comparing two comparators
Outcomes	• Cystometric first sensation and bladder capacity • Cystoscopic appearance • Maximal bladder capacity (hydrodistension) • Mast cell count • Voided urine volume • Urinary frequency • Global Response Assessment (GRA) • Pain Visual Analogue Scale (VAS) • O'Leary-Sant (OLS) Interstitial Cystitis Symptom Index (ICSI) and Interstitial Cystitis Problem Index (ICPI) • Pelvic Pain and Urgency/Frequency Symptom Scale (PUF) • Patient-reported improvement and pain scales	Outcome not listed in the inclusion criteria
Study type	Randomised controlled trials (RCTs)	• Reviews/systematic reviews/pooled trial analyses • Studies indexed as case reports, case series, editorials and letters • Conference abstracts • Non-human studies

Appendix D of the CS[1] reports that the citation sifting stage and study selection at the full-text stage were undertaken by two reviewers, which is considered best practice in systematic reviewing. However, it is not clear if, at both of these stages of the study selection process, the reviewers worked collaboratively or independently (the latter reflects best practice). It is also not clear in the CS (CS, Appendix D)[1] what proportion of citations at the sifting stage were double-checked (i.e., by both reviewers).

4.1.3 Critique of data extraction

Details regarding the company's data extraction methods (number of reviewers involved, items extracted, or a copy of a data extraction sheet) are not reported in the CS.[1]

Data extracted from the four included PPS in IC/BPS RCTs[14-17] are reported in Sections 4.1.4 to 4.2 and data extracted from the two Uracyst® in BPS RCTs[20, 21] reported in the CS[1] are reported below in Section 4.3. All data were checked against the published trial reports[14-17, 20, 21] by the ERG. Although the CS reports that two reviewers were involved in the study selection process, it is unclear how many were involved in the data extraction process and the ERG identified several data extraction errors. However, these errors did not impact on the analyses undertaken by the company.

4.1.4 Quality assessment

Quality assessment of the four RCTs of PPS in IC/BPS[14-17] is presented in Section B.2.5 and Appendix D of the CS.[1] The CS does not report where the quality assessment items were taken from, only that these were 'NICE criteria'. The ERG sought clarification with the company regarding this issue. The company's clarification response[1] stated that the items assessed were taken from the NICE Guidelines Manual.[23] These are appropriate criteria for assessing the methodological quality/risk of bias in RCTs.

It is considered good systematic review practice for two reviewers either to independently perform quality assessment or to check assessed items, but this was not reported in the CS. The ERG checked the company's quality assessment against the publications of the RCTs relevant to the decision problem.

Table 11 presents the company's quality assessment of the four RCTs of PPS in IC/BPS RCTs14-17 (Section 4.2.4 of this report).

4.1.5 Evidence synthesis

The company presented a narrative synthesis of the evidence for PPS in IC/BPS and sodium chondroitin sulphate instillations (Uracyst®) in BPS. The ERG considers the narrative synthesis

approach undertaken by the company to be acceptable. In addition, the company provided the following justification for not undertaking a network meta-analysis (CS, page 82): "*Twelve trials met the inclusion criteria. Six trials compared PPS capsules to oral placebo, three Uracyst® to placebo instillation and one each of Uracyst® to DMSO instillation, iAluRil® to DMSO instillation and Cystistat® to Gepan® . It was therefore not possible to construct a network comparing PPS to all relevant comparators. Only one bladder instillation, Uracyst®, could potentially be compared to PPS indirectly via placebo. However, there was considerable heterogeneity in the trials, which would make a robust ITC of PPS with any comparator challenging.*"

The company undertook a pairwise meta-analysis of RCTs of PPS compared to placebo in IC/BPS, a pairwise meta-analysis of Uracyst® compared to placebo in BPS, and an ITC of PPS in IC/BPS compared to Uracyst® in BPS. Further details of the PPS trials can be found in Section 4.2, further details of the Uracyst® trials can be found in Section 4.3, and further details of the ITC can be found in Section 4.4.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

4.2.1 Included trials of PPS in IC/BPS

The company identified four RCTs of PPS which were considered relevant to the decision problem (Sant *et al.*, 2003;[17] Parsons *et al.*, 1993;[15] Mulholland *et al.*, 1990;[14] Parsons and Mulholland, 1987[16]). All four trials included a comparison for PPS 100 mg three times per day to placebo. Sant *et al.*, (2003)[17] also evaluated hydroxyzine 50 mg administered orally once daily and hydroxyzine plus PPS (four treatment groups) in a factorial design. Placebo was three times per day.

In the RCT reported by Parsons and Mulholland (1987),[16] one study centre compared PPS 100 mg three times per day to PBO, and one study centre compared PPS 200 mg twice per day to PBO (two study centres). However, results for this trial are presented as both PPS groups combined compared to both PBO groups combined. The CS[1] (page 38) reports that 200 mg twice per day is comparable to the approved dose (300 mg per day). The EMA (EPAR, page 58) reports that across the pivotal studies, few patients received a dose of PPS 200 mg twice per day. However, the associated patient numbers are not presented.[8]

Eligibility criteria

All four RCTs recruited patients age ≥18 years old. With reference to the decision problem criteria in the NICE scope for IC/BPS patients with Hunner's lesions and/or glomerulations: in Sant *et al.* (2003)[17], IC/BPS was confirmed by cystoscopy and hydrodistention, following NIDDK[4] and Digestive and Kidney Diseases criteria;[24] in Mulholland *et al.* (1990)[14], patients had to have

cystoscopic examination under anaesthesia showing petechial haemorrhages or ulcers; and in Parsons and Mulholland (1987)[16], patients also had to have cystoscopic examination showing ulcer or petechial haemorrhage. In Parsons *et al*. (1993),[15] patients were recruited based on bladder capacity, number of voids per day, voided volume, and nocturia. Patients lacking one or two of these criteria had to also have pain and/or moderate urgency, negative urinary cytology studies and cultures, and cystoscopic findings of petechial haemorrhages and blood in the fluid return after bladder distension.

Eligibility criteria the four PPS in IC/BPS RCTs included in the CS are presented in Table 2.

Table 2: Patient eligibility criteria for the pentosan polysulfate sodium RCTs relevant to the decision problem (adapted from Tables 10 to 13 of the CS)

Trial	Inclusion criteria	Exclusion criteria
Sant *et al.*, 2003[17]	Patients ≥18 years old with a diagnosis of IC/BPS, confirmed by cystoscopy and hydrodistention, following National Institutes of Health-National Institute of Diabetes and Digestive and Kidney Diseases criteria. Patients had moderate symptoms of urinary frequency (≥11 time/day) and pain/discomfort (≥4 on a 0–9 Likert scale) for >24 weeks prior to trial entry.	Patients with history of: cyclophosphamide, pelvic radiation, augmentation cystoplasty, cystectomy, or cystolysis, neurectomy, implanted peripheral nerve stimulator, prostate surgery or treatment (men only). In prior 24 weeks: intravesical bacillus Calmette-Guerin, cystocele, rectocele, urinary incontinence surgery, transvaginal surgery, hysterectomy, prolapse, vaginal delivery, or caesarean section (women only) Prior 6–12 weeks: urethral dilatation, cystometrogram, urodynamics, cystoscopy/hydrodistention, bladder biopsy, prostate biopsy (men only), any intravesical treatment other than BCG Prior 4 weeks: initiation of any new medications for IC, washout for oral PPS and hydroxyzine Any history of bladder calculus, tuberculous cystitis, neurological disease or diabetic cystopathy, malignant bladder tumours, urethral cancer Last 3 years: uterine, cervical or vaginal cancer (women only) Last 6–12 weeks: bacterial urinary tract infection; active genital herpes, gross haematuria Concurrent: active urethral calculus, ureteral calculus, symptomatic urethral diverticulum, documented chronic bacterial prostatitis (men only), active vaginitis, pregnant, breast-feeding (women only) Concurrent: urinary void with a maximum volume >350 cc; residual urine volume ≥150 cc by ultrasound or catheter (men only), liver function test >1.5× upper limit of normal, abnormal blood coagulation tests
Parsons *et al.*, 1993[15]	Patients ≥18 years old with 8 or more voids per day; average voided volume of 50–200 cc; anaesthetic bladder capacity of 350–1,000 cc; and nocturia (at least 1 or 2 episodes) OR any patients lacking 1 or 2 of these criteria if they had pain and/or moderate urgency, negative urinary cytology studies and cultures, and cystoscopic findings of petechial haemorrhages and blood in the fluid return after bladder distension	Patients <18 years old or who were unavailable for the duration of the trial or unable to follow instructions; pregnant or lactating women; premenopausal women not practicing an effective means of birth control. Patients with evidence of active bleeding peptic ulcer disease or bleeding diathesis; signs of recurrent bacteriuria or obvious neurological impairment. Patients who had: received previous treatment with known bladder irritants; a history of pelvic irradiation, bladder carcinoma, urinary tuberculosis or schistosomiasis; a known allergy to PPS

Trial	Inclusion criteria	Exclusion criteria
Mulholland *et al.*, 1990[14]	Patients with urgency expressed as moderate on a 5-point analogue scale (not reported in the trial report if it is a visual analogue scale or not); frequency of at least 10 voids/day; nocturia of at least 2 voids/night; pain as recorded on a 5-point analogue scale; continuous duration of symptoms of at least 1 year; failed previous conventional therapy e.g., chlorpactin, hydrodilatation, DMSO; average voided volume of 200 ml or less measured over a 3-day period; negative urine culture and cytology; cystoscopic examination under anaesthesia showing petechial haemorrhages or ulcers with gross blood in the fluid return and a bladder capacity of 800 ml or less	Patients aged <18 years; lack of availability for the duration of the trial or inability to follow instructions; pregnancy; premenopausal and not practicing effective means of birth control; lactating mothers; evidence of active bleeding peptic ulcer disease; bleeding diathesis; known allergy to PPS; treatment with PPS within six weeks of trial; signs of: recurrent bacteriuria, obvious neurologic impairment, history of pelvic irradiation, previous treatment with known bladder irritants, bladder carcinoma, urinary tuberculosis, shistosomiasis
Parsons and Mulholland, 1987[16]	Patients aged >18 years old with ≥1 year of symptoms (urgency, frequency, nocturia and/or pain), negative urine cultures, cystoscopic examination showing ulcer or petechial haemorrhage (after bladder distension), biopsy-proved inflammation, and negative cytology studies.	Not reported

Trial characteristics

Details of trial location treatments and numbers randomised, prohibited concomitant medications and other outcomes reported by the four PPS in IC/BPS RCTs included in the CS are presented in Table 3.

All four RCTs of PPS in IC/BPS were multicentre trials conducted in the USA.[14-17] The number of centres ranged from two[16] to seven.[15, 17] Numbers randomised to PBO and PPS 100 mg were 31 and 29 respectively in Sant *et al.* (2003),[17] 74 and 74 respectively in Parsons *et al.* (1993),[15] and 56 and 54 respectively in Mulholland *et al.* (1990).[14] Parsons and Mulholland (1987)[16] did not report numbers randomised by group, but that 75 patients were randomised across two centres to PBO, PPS 100 mg, or PPS 200 mg.

In Sant *et al.* (2003),[17] prohibited medication included: cimetidine, intravesical heparin, chronic use of acetylsalicylic acid, nonsteroidal anti-inflammatory drugs, or sedating histamine-1 receptor antagonists. Prohibited medications were similar in Parsons *et al.* (1993)[15] and Mulholland *et al.*, 1990.[14] The RCT by Parsons and Mulholland, 1987[16] did not report on permitted or prohibited medication.

Global response assessment (GRA) varied across the four RCTs of PPS in IC/BPS.[14-17] In Sant *et al.* (2003),[17] responders were those who six or seven (moderately or markedly improved) on a seven-point scale (markedly worse, moderately worse, slightly worse, no change, slightly improved, moderately improved and markedly improved). In Parsons *et al.* (1993),[15] responders were those with >50% overall improvement in symptoms (improvement rated as: slight, 25%; moderate, 50%; great, 75%; symptoms gone, 100%). In Mulholland *et al.* (1990),[14] a >50% overall improvement in symptoms on a six-point scale ranging from worse to excellent was considered by the company as comparable to GRA for the purpose of analysis. In Parsons and Mulholland (1987),[16] symptoms of urgency, frequency, nocturia and pain were graded as 0%, 25%, 50%, 75%, or 100% improvement. The company considered >50% pain improvement comparable to GRA for the purpose of analysis. The ERG's clinical advisors did not all agree that the measures of GRA were comparable across the RCTs.

One RCT reported that outcome follow-up was at 24 weeks,[17] and two reported that outcome follow-up was at three months.[14, 15] In the RCT by Parsons and Mulholland (1987),[16] if the patient failed to respond to therapy at three months (PPS or PBO), patients were switched to the alternative treatment (from PPS to PBO, or from PBO to PPS). The CS[1] reports data at three months, prior to the switch.

Table 3: Trial locations, treatments and numbers randomised, concomitant medication, and outcomes for the pentosan polysulfate sodium RCTs relevant to the decision problem (adapted from Tables 10 to 13 of the CS)

Trial Location	Treatments, numbers randomised and follow-up	Permitted and prohibited concomitant medication	Primary outcomes	Other outcomes used in the economic model/specified in the scope
Sant *et al.*, 2003[17] **USA (7 centres)**	PBO, 31 PPS 100 mg, 29 Both TID	Prohibited: cimetidine, intravesical heparin, chronic use of acetylsalicylic acid, nonsteroidal anti-inflammatory drugs, or sedating histamine-1 receptor antagonists	*Global Response Assessment:* - n (%) moderately/markedly improved (score of 6 or 7) (24 weeks)	*Pain and urgency (scope only):* - mean change in non-VAS pain score (time point: 24 weeks) - mean change in urgency score (time point: 24 weeks) *O'Leary-Sant/ICSI and ICPI scores (model and scope):* - mean ICSI change (time point: 24 weeks) - mean ICPI change (time point: 24 weeks) *Urinary frequency (scope only):* - - mean daily frequency change (time point: 24 weeks)
Parsons *et al.*, 1993[15] **USA (7 centres)**	PBO, 74 PPS 100 mg, 74 Both TID	Prohibited: anticoagulant therapy; chronic use of narcotics; artificial sweeteners; PPS within 4 weeks of the trial	*Global Response Assessment:* - n patients reporting >50% overall improvement in symptoms (time point: 3 months)	*Voided urine volume (scope only):* - mean volume/void change (cc) (time point: 3 months) - % patients with increase of >20 cc in volume/void (time point: 3 months) - mean total daily volume change (cc) (time point: 3 months) *Pain and urgency (scope only):* - patient-reported degree of pain and urgency on a scale of 0 to 5, in which 0 is none, 1 is mild, 3 is moderate, and 5 is severe (time point: 3 months) *Investigator evaluation of overall*

Trial Location	Treatments, numbers randomised and follow-up	Permitted and prohibited concomitant medication	Primary outcomes	Other outcomes used in the economic model/specified in the scope
				improvement (scope only): - - overall changes in condition were evaluated as worse, no change, fair (25%), good (50%), very good (75%), and excellent (100%) (time point: 3 months)
Mulholland *et al.*, 1990[14] **USA (5 centres)**	PBO, 56 PPS 100 mg, 54 Both TID	Prohibited: anticoagulant therapy; chronic use of narcotics; use of artificial sweeteners; treatment within PPS within 6 weeks of the trial	*6-point patient-reported improvement (considered comparable to GRA for the purpose of analysis):* - n patients reporting >50% overall improvement in symptoms (time point: 3 months)	*6-point investigator-evaluated improvement (scope only):* - % >50% improved (time point: 3 months) *Patient-reported pain improvement (scope only):* - % >50% improved (time point: 3 months) - % reporting decrease of >1 point (time point: 3 months) - mean reduction in pain score (time point: 3 months) *Voided urine volume (scope only):* - Mean volume/void change (cc) (time point: 3 months) - % patients with increase of >20 cc in volume/void (time point: 3 months) - Mean total daily volume change (cc) (time point: 3 months)
Parsons and Mulholland, 1987[16] **USA (2 centres)**	Total PBO, PPS 100 mg TID, and PPS 200 mg BID; 75 For 3 months initially then, if PBO or PPS failure, cross-over to PBO	NR	*Global Response Assessment:* - Patient-reported pain improvement (time point: 3 months [before crossover]) (considered comparable to GRA for the purpose of analysis) *Urinary frequency:* - n (%) any improvement (time point:	*Voided urine volume (scope only):* - mean volume/void (mL) (time point: 3 months [before crossover])

Trial Location	Treatments, numbers randomised and follow-up	Permitted and prohibited concomitant medication	Primary outcomes	Other outcomes used in the economic model/specified in the scope
	or PPS for a further 3 months		3 months [before crossover]) - mean daily change (improved patients only) (time point: 3 months [before crossover]) - mean daily frequency (time point: 3 months [before crossover]) *Urinary urgency:* - n (%) any improvement (time point: 3 months [before crossover]) - mean % improvement (time point: 3 months [before crossover]) *Nocturia:* - mean improvement (time point: 3 months [before crossover])	

BID, twice per day; GRA, global response assessment; ICPI, Interstitial Cystitis Problem Index; ICSI, Interstitial Cystitis Symptom Index; NR, not reported; PBO, placebo; PPS, pentosan polysulfate sodium; TID, three times per day

Sample size and power calculation

The CS[1] (Table 17) reports on sample size and power calculations. In the Sant *et al.* (2003) RCT,[17] "*the projected sample size of 136 participants planned to be recruited during 10 months was selected to detect a difference in response rates of 30% and 65% (80% power at a 2-sided significance level of 5%)*". One hundred twenty-one (121) participants were randomised to four treatment groups. No sample sizes were defined prospectively for the trials by Parsons *et al.* (1993);[15] Mulholland *et al.* (1990);[14] or Parsons and Mulholland (1987);[16]. The European Medicines Agency (EMA) consider this is a weakness of the literature-based application in the European public assessment report (EPAR).[8]

Baseline characteristics of trial participants

Details of participant baseline characteristics in the four PPS in IC/BPS RCTs included in the CS are presented in Table 4.

The proportion of patients who were female across treatment groups was >89% in all four RCTs.[14-17] Where age was reported, patients were in the fifth decade of life.[14, 15, 17]

The RCTs by Parsons *et al.* (1993)[15] and Mulholland *et al.* (1990)[14] reported Hunner's ulcers in ≤8% of patients. However, Parsons and Mulholland (1987) reported that across PBO, PPS 100 mg, and PPS 200 mg treatment groups, 28% had Hunner's ulcers. Sant *et al.* (2003)[17] did not report on Hunner's ulcers. Petechial haemorrhage was not reported by Sant *et al.* (2003)[17] or Parsons and Mulholland (1987). Across the RCTs, Parsons *et al.* (1993)[15] and Parsons and Mulholland (1987),[16] the proportions of patients with petechial haemorrhage varied depending on numbers with haemorrhages, but with between 40% and 50% of patents having a moderate number (not defined) of petechial haemorrhage in both of these RCTs.

The ERG's clinical advisors believed that the populations in these RCTs were generally comparable to the UK IC/BPS population.

Baseline pain and urinary details were only reported by Sant *et al.* (2003),[17] and baseline bladder capacity was only reported by Parsons *et al.* (1993)[15] and Mulholland *et al.* (1990).[14]

Table 4: Baseline characteristics of participants in the pentosan polysulfate sodium RCTs relevant to the decision problem (adapted from Table 16 of the CS and the trial reports)

Trial Location	n/N (%) female	Mean (SD) age years	N/n (%) with ulcers/haemorrhage	Other characteristics
Sant *et al.*, 2003[17]	PBO, 28/31 (90%) PPS, 26/29 (90%)	PBO, 41.6 (15.5) PPS, 48.7 (15.1)	NR	*Prior symptoms for ≥ 52 weeks, n (%):* PBO, 28 (90%); PPS, 28 (96%) *Pain score (0 to 9), mean (SD):* PBO, 6.0 (1.3); PPS, 6.3 (1.4) *Urinary score (0 to 9), mean (SD)* PBO, 6.5 (1.5); PPS, 6.9 (1.2) *24-hour frequency score (0 to 9), mean (SD):* PBO, 18.9 (10.3); PPS, 18.3 (6.8) *ICSI, mean (SD):* PBO, 14.6 (3.3); PPS, 14.3 (3.3) *ICPI, mean (SD):* PBO, 12.8 (2.4); PPS, 12.8 (2.7) *Wisconsin IC score (0 to 42), mean (SD):* PBO, 32.9 (6.7); PPS, 30.4 (6.8)
Parsons *et al.*, 1993[15]	PBO, 74/74 (100%) PPS, 66/74 (93%)	PBO, 45.5 (NR) PPS, 42.7 (NR)	Hunner's ulcer: PBO, NR (4%) PPS, NR (4%) Petechial haemorrhage: PBO, NR (none, 1%; few, 8%; moderate, 43%; many, 47%) PPS, NR (none, 1%; few, 9%; moderate, 41%; many, 49%)	*Other abnormalities:* PBO, 8%; PPS, 11% *Bladder capacity under anaesthesia, mean (cc):* PBO, 601; PPS, 656
Mulholland *et al.*, 1990[14]	PBO, 45/56 (87%) PPS, 49/54 (91%)	PBO, 45.3 (NR) PPS, 43.3 (NR)	Hunner's ulcer: PBO, NR (4%) PPS, NR (8%) Petechial haemorrhage: PBO, NR (few, 27%; moderate, 48%; many, 25%) PPS, NR (few, 26%; moderate, 46%; many, 28%)	*Disease duration mean years:* PBO, 5.6; PBO, 7.4 *Other abnormalities:* PBO, 11%; PPS, 4% *Bladder capacity under anaesthesia, mean (cc):* PBO, 585; PPS, 569 *Patients with severe disease:* PBO, 59%; PPS, 59%

Trial Location	n/N (%) female	Mean (SD) age years	N/n (%) with ulcers/haemorrhage	Other characteristics
Parsons and Mulholland, 1987[16]	Overall (PBO, PPS 100 mg & 200mg), 68/75 (90%)	NR	Ulcers: Overall, 28% Haemorrhage: NR	NR

NR, not reported; PBO, placebo; PPS, pentosan polysulfate sodium

Participants completing / included in analysis in PPS RCTs

In the RCT by Sant *et al.* (2003),[17] an intention-to-treat analysis for the primary endpoint of GRA was used, where all participants who did not complete the 24-week follow-up assessment were classified as non-responders. In the RCT by Parsons *et al.* (1993),[15] 148 participants were randomised and the proportion of participants with the primary endpoint for 50% overall improvement is expressed as a proportion of the number randomised per group (n=74). However, nine participants per group were reported as not completing the study.

In the RCT by Mulholland *et al.* (1990),[14] whilst 110 participants were randomised, it is unclear from the trial report how many patients contributed data to each of the analyses as only the proportion (%) of participants (not n/N) with overall improvement at three months and other outcomes are reported. Three participants treated with PPS and nine treated with PBO failed to complete the study. However, the CS[1] reports that the primary efficacy analysis was as intention-to-treat (all participants randomised) (CS, Figure 31).

In the RCT by Parsons and Mulholland (1987),[16] 62 of the 75 participants randomised were reported to have completed the study, which included two study phases - before and after treatment switching (from PPS to PBO, or from PBO to PPS at three months). The CS[1] reports data at three months, prior to the switch. The CS[1] reports that the primary efficacy analysis was based on completers (n=62) (CS, Figure 33). However, the numbers in the trial report prior to switching (Tables 1 and 5 of the trial report) are discrepant with this. The ERG also notes that participant numbers after switching (Table 2 of the trial report) are greater than the number randomised, implying that double-counting of patients might have occurred in the analyses following cross-over.

Trial authorship

The ERG notes that there is some author commonality across all four RCTs of PPS in IC/BPS. The author Parsons is cited as a trial author on three of the trial reports,[14-16] the author Mulholland is cited as an author on two trial reports,[14, 16] and the author Sant is cited as an author on three trial reports.[14, 15, 17] All four RCTs were undertaken in the USA and published between 1987 and 2003. The ERG notes there have been no other published independent studies validating the results of these RCTs.

The Food and Drug Administration (FDA)[25] statistical and medical reviews note that, as part of the 1994 non-approval issues, that the RCTs by Mulholland *et al.* (1990),[14] and Parsons *et al.* (1993)[15] were not considered to be independent because the majority of the efficacy database for each of these studies was generated by the same three site investigators. One of the Medical Officers for the FDA observed that three investigators (Hanno, Parsons, and Sant) participated in both of the RCTs by Mulholland *et al.* (1990),[14] and Parsons *et al.* (1993),[15] and that these three investigators were

accountable for 75% (82 of 110) patients in Mulholland *et al.* (1990),[14] and 57% (95 of 148) patients in Parsons *et al.* (1993).[15] As such, that these RCTs could not be considered as independent trials.[25]

The FDA also notes that the RCT by Parsons *et al.* (1993),[15] may have included a positive "treatment-by-investigator effect" for one of the seven included study sites. When data from the site were excluded from the analysis, a trend in favour of PPS remained, but was no longer statistically significant.[25] The FDA notes regarding the site investigator that the sponsor submission states that (page 258): *"Dr Parsons had a prior arrangement with [FDA redacted] to receive a royalty on the sales of Elmiron"* (FDA page 258).[25]

4.2.2 Efficacy results for trials of PPS in IC/BPS

Global response assessment

Details of the three RCTs of PPS in IC/BPS that reported GRA as an outcome[14, 15, 17] are presented in Table 5. In the RCT by Sant *et al.* (2003),[17] which used a factorial design, a greater proportion of patients receiving PPS (PPS and PPS plus hydroxyzine groups combined) had a GRA score of six or seven compared to PBO (PPS and hydroxyzine placebo groups combined) at 24 weeks, but the between-group difference in proportions was not statistically significant (PBO 18% vs. PPS 34%, p=0.064, CI not reported).

The trials by Parsons *et al.* (1993)[15] and Mulholland *et al.* (1990)[14] both reported the proportions of patients with a >50% improvement in GRA as both patient-reported and investigator-reported outcomes. The between-group difference in patient-reported GRA at three months was statistically significant in favour of PPS in both the Parsons *et al.* (1993) trial (5-point scale, PBO 16% vs PPS 32%, p=0.01, CI not reported)[15] and the Mulholland *et al.* (1990) trial (6-point scale, PBO 13% vs PPS 28%, p=0.04, CI not reported).[14] The between-group difference investigator-reported GRA at three months was also statistically significant in favour of PPS in both the Parsons *et al.* (1993) trial (5-point scale, PBO 15% vs PPS 36%, p=0.002, CI not reported)[15] and the Mulholland *et al.* (1990) trial (6-point scale, PBO 11% vs PPS 26%, p=0.03, CI not reported).[14]

The CS[1] reports that the GRA assessment methods in Parsons *et al.* (1993)[15] and Mulholland *et al.* (1990)[14] were considered by the company to be equivalent to GRA scored as six or seven on a seven-point scale, as this was considered equivalent by the EMA[26] (CS, page 62).

The ERG notes that the Sant *et al.* (2003) trial[17] was a feasibility study that reported that a prospective Phase 3 study was not warranted. The authors report that the reason for this was partly because the investigators concluded that PPS did not improve the GRA sufficiently to initiate a larger clinical trial in spite of the authors stating that a minimal important clinical difference had not been determined by

the trial, and not giving consideration to the range of plausible treatment effects that would be suggested by confidence intervals (no CIs were reported).[17] In addition, the CS reports a "further analysis" of GRA that suggested that the effect of PPS was statistically significant (p=0.039) (CS, Section B.2.8.1),[1] whereas Sant *et al.* (2003)[17] reported the *p*-value as 0.064 (CI not reported). The difference between these two *p*-values seems to be because Sant *et al.* (2003)[17] accounted for clinical centre clustering using a Mantel-Haenzsel test, whereas the CS ignored clustering and used a Z-test. Furthermore, the distinction is important when considering the meta-analysis using the evidence from Sant *et al.* (2003)[17] because the company's approach effectively underestimates the standard error of the sample estimate of treatment effect.

Table 5: Details of global response assessment in the pentosan polysulfate sodium RCTs in IC/BPS (adapted from the CS Table 19)

Trial	**Sant *et al.*, 2003[17]**	**Parsons *et al.*, 1993[15]**	**Mulholland *et al.*, 1990[14]**
GRA assessment method	Score of 6-7 on 7-point scale	>50% overall improvement in symptoms on a 5-point scale	>50% overall improvement in symptoms on a 6-point scale
Follow-up time point	24 weeks	3 months	3 months
N (%) score of 6 or 7	PBO, 11/62 (18) PPS 20/59 (34)	NR	NR
P value (between groups)	0.064	NA	NA
N (%) ≥50% improved (patient-reported)	NR	PBO, 12/74 (16) PPS, 24/74 (32)	PBO, NR (13) PPS, NR (28)
P value (between groups)	NA	0.01	0.04
N (%) ≥50% improved (investigator-reported)	NR	PBO, NR (15) PPS, NR (36)	PBO, NR (11) PPS, NR (26)
P value (between groups)	NA	0.002	0.03

NA, not applicable; NR, not reported; PBO, placebo; PPS, pentosan polysulfate sodium

Pain data from Parsons and Mulholland 1987 used as a proxy for GRA in the CS analyses

For outcome data, please see the next section on non-VAS pain outcomes in this ERG report.

Although the pain data presented in the CS for Parsons and Mulholland (1987)[16] at three months concur with the trial report[16] (PBO, 3/20 (15%); PPS, 12/27 (44%); CS Table 21), these data do not concur with the data for this RCT presented in the GRA forest plot in Figure 11 of the CS (PBO, 6/37; PPS, 15/38).[1] However, the data in the CS Figure 11 for Parsons and Mulholland (1987)[16] do concur with those presented by the EMA in the EPAR (EPAR, Table 30).[8] The EPAR states (EPAR, page 91): *"Although no global response assessment was conducted in the study reported by Parsons and Mulholland, 1987, the data imputation used for the meta-analysis conducted by the applicant is deemed sufficiently comparable."* However, details of this data imputation are not reported in the EPAR.[8]

Details of non-VAS pain outcomes for all four RCTs of PPS in IC/BPS[14-17] are presented in Table 6 of this ERG report.

Non-VAS pain outcomes

All four RCTs of PPS in IC/BPS reported on non-VAS pain,[14-17] assessment of this outcome varied. Details of the assessment methods and results are presented in Table 6. Sant *et al.* (2003)[17] used a patient-reported 0–9 Likert scale (lower is better, participant inclusion criterion score of ≥4), Parsons *et al.* (1993)[15] and Mulholland *et al.* (1990)[14] both assessed pain on a 0–5 scale (0 = no pain, 5 = severe pain). The RCT by Parsons and Mulholland (1987)[16] assessed patient-graded improvements of 0%, 25%, 50%, 75%, or 100%.

Between-group differences in change-from-baseline were reported by Sant *et al.* (2003)[17] at 24 weeks and Mulholland *et al.* (1990)[14] at three months. Both reported a reduction in change-from-baseline in both PPS and PBO. Sant *et al.* (2003)[17] reported PPS -0.8 vs. PBO -1.0 and Mulholland *et al.* (1990)[14] reported PPS -0.05 vs. PBO -0.02 (incorrectly reported in the CS as PPS 0.05 vs. PBO 0.02). In both trials, the between-group difference was not statistically significant (*p*-values or CIs not reported).

Parsons *et al.* (1993),[15] Mulholland *et al.* (1990)[14] and Parsons and Mulholland (1987),[16] all reported on the proportion of participants with a >50% pain improvement at three months. Respective values were: PPS 18% vs. PBO 38% (p=0.005), PPS 27% vs, PBO 14% (p=0.08), and PPS 44% vs. PBO 15% (p=0.02, CI not reported).

Parsons *et al.* (1993)[15] and Mulholland *et al.* (1990),[14] also reported on the proportion of participants with a decrease of >1 point at three months. Respective values were: PPS 66% vs. PBO 51% (p=0.04 in trial report, CI not reported;[15] incorrectly reported in CS as p=0.004), and PPS 46% vs. PBO 29% (p=0.07, CI not reported).

Parsons and Mulholland (1987),[16] also reported on the mean percentage improvement at three months: PPS 33.3 (SD 35) vs. PBO 12.2 (SD 14.3) (p=0.02, CI not reported).

Table 6: Details of non-VAS pain outcomes in the pentosan polysulfate sodium RCTs in IC/BPS (adapted from the CS Table 21)

Trial	Sant *et al.*, 2003[17]	Parsons *et al.*, 1993[15]	Mulholland *et al.*, 1990[14]	Parsons and Mulholland, 1987[16]
Follow-up time point	24 weeks	3 months	3 months	3 months
Pain measurement scale	PR: 0–9 scale	PR: 0–5 scale	PR: 0–5 scale	PR: 0%, 25%, 50%, 75% or 100% improvement
Mean (SD) score (baseline)	PBO, 6.0 (1.3) PPS, 6.3 (1.4)	NR	NR	NR
Mean (SD) score (follow-up)	NR	NR	NR	NR
Mean (SD) change from baseline	PBO, -1.0 (1.8) PPS, -0.8 (1.8)	NR	PBO, -0.02 (NR) PPS, -0.05 (NR)	NR
P value (change from baseline)	NR	NA	PBO, NS/NR PPS, 0.05	NA
P value (between groups)	NS	NA	NS	NA
N (%) >50% improved	NR	PBO, NR (18%) PPS, NR (38%)	PBO, NR (14%) PPS, NR (27%)	PBO, 3/20 (15%) PPS, 12/27 (44%)
P value (between groups)	NA	0.005	0.08	0.02
N (%) decrease of >1 point	NR	PBO, NR (51%) PPS, NR (66%)	PBO, NR (29%) PPS, NR (46%)	NR
P value (between groups)	NA	0.04	0.07	NA
Mean (SD) % improvement	NR	NR	NR	PBO, 12.2 (14.3) PPS, 33.3 (35)
P value (between groups)	NA	NA	NA	0.02

NA, not applicable; NR, not reported; NS, not significant; PBO, placebo; PPS, pentosan polysulfate sodium; PR, patient-reported; SD, standard deviation

O'Leary-Sant Interstitial Cystitis Symptom Index and Problem Index scores

The RCT by Sant *et al.* (2003),[17] was the only RCT of PPS in IC/BPS in the CS to report on Interstitial Cystitis Symptom Index and Problem Index scores (ICSI and ICPI).[22] In Sant *et al.* (2003),[17] which used a factorial design, there was no statistically significant between-group difference in change over time in either ICSI or ICPI (*p*-values or CIs, not reported). Details of these outcomes are presented in Table 7.

Table 7: Details of O'Leary-Sant Interstitial Cystitis Symptom and Problems for Sant *et al.* (2003) (adapted from the CS Table 19)

Trial	**Sant *et al.*, 2003[17]**
Follow-up time point	24 weeks
Mean (SD) ICSI score (baseline)	PBO, 14.6 + 3.3 PPS 14.3 + 3.3
Mean (SD) ICSI score change from baseline	PBO, -1.7 (3.5) PPS -2.6 (3.4)
P value (between groups)	NS
Mean (SD) ICPI score (baseline)	PBO, 12.8 + 2.4 PPS 12.8 + 2.7
Mean (SD) ICPI score change from baseline	PBO, -1.9 (2.8) PPS -2.6 (3.5)
P value (between groups)	NS
NS, not significant; PBO, placebo; PPS, pentosan polysulfate sodium; SD, standard deviation	

Daily urinary frequency

Two RCTs of PPS in IC/BPS, assessed daily urinary frequency.[16, 17] Details of the assessment methods and results are presented in Table 8.

At 24 weeks, Sant *et al.* (2003)[17] that there was no statistically significant between-group difference in change-from-baseline in mean daily frequency (PBO, -0.5 (SD 5.3) vs. PPS, -0.2 (SD 5.0); *p*-value or CI, not reported).

In the RCT by Parsons and Mulholland (1987),[16] at three months there were no statistically significant between-group differences evident in the proportion of participants with any improvement at follow-up (PBO, 10/24 (42%) vs. PPS, 20/31 (65%); $p=0.06$, CI not reported), or mean change from baseline in frequency (PBO, -1.8 vs. PPS, -5.4; $p=0.06$; SDs or CIs, NR).

Table 8: Details of daily urinary frequency outcomes in the pentosan polysulfate sodium RCTs in IC/BPS (adapted from the CS Table 22)

Trial	Sant *et al.*, 2003[17]	Parsons and Mulholland, 1987[16]
Follow-up time point	24 weeks	3 months
Mean (SD) daily frequency (baseline)	PBO, 18.9 (10.3) PPS, 18.3 (6.8)	PBO, 18.8 (NR)* PPS, 18.0 (NR)
Mean (SD) daily frequency (follow-up)	NR	PBO, 19.5 (NR) PPS, 18.0 (NR)
Mean (SD) daily frequency (change from baseline)	PBO, -0.5 (5.3) PPS, -0.2 (5.0)	NR NR
P value (between groups)	NS	*P*=0.06
N (%) any improvement (follow-up)	NR	PBO, 10/24 (42%) PPS, 20/31 (65%)
P value (between groups)	NA	*p*=0.06
Mean change (improved patients, change from baseline)	NR	PBO, -1.8 PPS, -5.4
P value (between groups)	NA	*p*=0.06

NA, not applicable; NR, not reported; NS, not significant; PBO, placebo; PPS, pentosan polysulfate sodium; SD, standard deviation
* Incorrect in CS, PBO reported as 18.0 in the CS

Volume/void outcomes

Three of the RCTs of PPS in IC/BPS, assessed volume/void outcomes.[14-16] Details of the assessment methods and results are presented in Table 9.

In the RCTs by Parsons *et al.* (1993)[15] and Mulholland *et al.* (1987),[16] at three months there were no statistically significant between-group difference evident in the mean void volume (mL) at follow-up (PBO, -2.1 vs. PPS, 20.4; *p*-value NR; SDs or CIs, NR and PBO, 7.6 vs. PPS, 9.8; *p*-value NR; SDs or CIs, NR; respectively).

In the RCT by Parsons and Mulholland (1987),[16] at three months the respective values were PBO, 74.3 vs. PPS, 106.9 (SDs or CIs, NR). Table 23 of the CS[1] reports that the between-group difference was statistically significant at *p*=0.009. However, this *p*-value is for the PPS group only after treatment switching (Table 3 of the trial report,[16] values PBO 84.6 (SD 53), *p*=0.05 vs. PPS 102.5 (SD 57), *p*=0.009). A *p*-value or CI for the between-group difference prior to switching, at three months, is not reported in the trial report (Table 5 of the trial report[16]).

Table 9: Details of daily void/volume outcomes in the pentosan polysulfate sodium RCTs in IC/BPS (adapted from the CS Table 23)

Trial	Parsons *et al.*, 1993[15]	Mulholland *et al.*, 1990[14]	Parsons and Mulholland, 1987[16]
Follow-up time point	3 months	3 months	3 months
Mean volume/void, mL (baseline)	NR	NR	PBO, 76.7 PPS, 93.8
Mean volume/void, mL (follow-up)	NR	NR	PBO, 74.3 PPS, 106.9
Mean volume/void, mL (change from baseline)	PBO, -2.1 PPS, 20.4	PBO, 7.6 PPS, 9.8	NR
P value (change from baseline)	NR	NR	PBO, 0.6 PPS, 0.06
P value (between groups)	NS	NS	NR
Mean total daily voided volume, mL (change from baseline)	PBO, -42 PPS, 3	PBO, -20 PPS, 60	NR
P value (between groups)	NS	NS	NA
% patients with >20 mL increase (follow-up)	PBO, 25% PPS, 40%	PBO, 20% PPS, 30%	NR
P value (between groups)	0.02	NS	NA

mL, millilitre; NA, not applicable; NR, not reported; PBO, placebo; PPS, pentosan polysulfate sodium
* Incorrect in CS, PBO reported as 0.3 in the CS

Nocturia

In Table 24 of the CS, the company reports that in the RCT by Parsons and Mulholland (1987)[16] at three months the mean improvement in nocturia was PBO -0.09 (SD 0.8) vs. PPS -2.1 (SD 2.2), p=0.05 (CI not reported). This is the only RCT in IC/BPS for which the company report nocturia data in the CS.[1] However, the trial reports by Mulholland *et al.* (1990)[14] and Parsons *et al.* (1993),[15] both report on this outcome.

Mulholland et al (1990)[14] reported that at three months there was no statistically significant between-group difference in change in nocturia PBO -0.5 vs. PPS -0.8, *p*-value or CI, NR). Parsons *et al.* (1993)[15] also reported that at three months, there was no statistically significant between group difference in nocturia (no data reported). In Parsons *et al.* (1993),[15] increase in nocturia was recorded as an adverse event. The numbers (%) of patients experiencing this AE were PBO 0 (0%) vs. PPS 1 (1.4%) (*p*-value or CI, NR). This AE for Parsons *et al.* (1993)[15] is not presented in the Section B.2.10. of the CS on AEs, Table 32,[1] as there was not >1 patient in either treatment group with this AE.

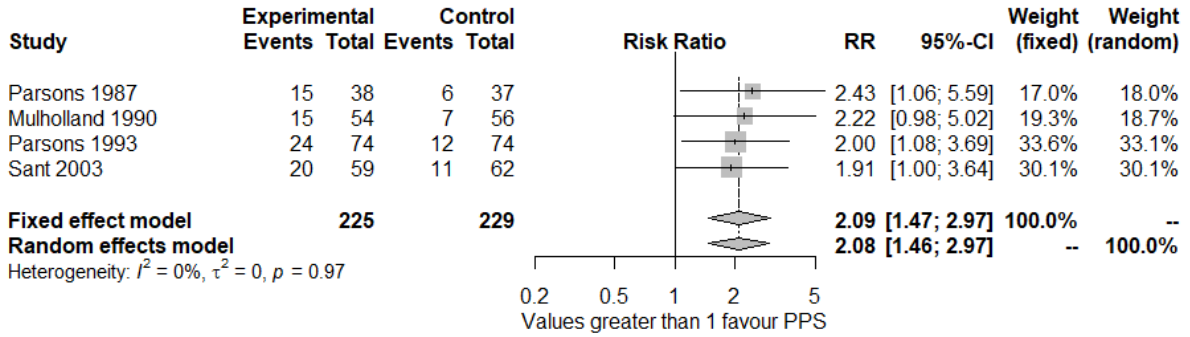
Other outcomes

No other clinical effectiveness outcomes for RCTs of PPS in IC/BPS were reported in the CS.[1]

Pairwise meta-analysis of effectiveness

The company presented a pairwise meta-analyses of GRA across the four RTCs of PPS in IC/BPS[14-17] (CS, figure 11). The forest plot for this analysis is presented in Figure 3 below. The fixed effect RR of 2.09 (95%CI: 1.47 to 2.97) was applied by the company in the economic model.

The pooled estimate was used by the company to compare to the pooled GRA estimate from the pairwise meta-analyses across the two Uracyst® RCTs[20, 21] based on the Bucher method. Further details of this are presented in Section 4.4 of this ERG report.



Reproduced from the CS page 84.[1]

Figure 3: Meta-analysis of Global Response Assessment in RCTs of PPS in IC/BPS (risk ratio) (CS, figure 11)

4.2.3 Safety results for trials of PPS in IC/BPS

Details of the summaries of adverse events presented in the CS for the four RCTs of PPS in IC/BPS[14-17] are presented in Table 10.

Sant *et al.* (2003),[17] which used a factorial design study, reported that there was no statistically significant difference in the overall adverse event rates between treatment arms (*p*-values or CIs, not reported). Parsons *et al.* (1993)[15] reported that there were no "*clinically significant differences between the treatment groups for any of the laboratory data, and there were no patients with laboratory findings critically outside the normal range for any of the parameters*" (*p*-values or CIs, not reported). Mulholland *et al.* (1990)[14] reported that the observed reactions were "*not different from those that might be observed in any random population over a three month period and were not serious*". Parsons and Mulholland (1987),[16] reported that among the 62 patients who completed the study, only a single side effect (skin rash) was noted in one participant. However, it was not reported which treatment group this was in, or whether this was before or after treatment switching at three months.

The CS[1] summarises that across the RCTs of PPS in IC/BPS, PPS is well tolerated. The ERG notes that AEs in Sant *et al.* (2003),[17] were recorded according to a grading system for 0 (none) to 3 (severe), and that >50% patients receiving PPS or PBO had moderate (grade 2) AEs. However, the trial report summarises that (page 812): "*The majority of [AEs] were minor, and not specifically related to PPS or hydroxyzine. The primary areas were constitutional symptoms (fatigue and drowsiness), gastrointestinal disturbances, and pain (abdominal/pelvic and other locations). There was no statistically significant difference in the overall adverse event rates between treatment arms*".[17]

Common adverse events (AEs) listed in the SmPC are: headache, dizziness, nausea, diarrhoea, dyspepsia, abdominal pain, abdominal enlargement, rectal haemorrhage, peripheral oedema, alopecia, back pain, asthenia, and pelvic pain.[8] Clinical advice received by the ERG from experience of using PPS on a named patient basis is that AEs are rare.

With respect to mortality, the EPAR for PPS reports that (EPAR, page 99): '*7 deaths were reported in a long-term, open-label study (Hanno et al. 1997),[27] considered as not related to study medication. 3 deaths were reported in the study published by Jepsen et al. (1998),[28] considered as not related to study medication.*'[8]

Table 10: Details of adverse events in the pentosan polysulfate sodium RCTs in IC/BPS (adapted from the CS Tables 31 to 34)

Sant *et al.*, 2003[17]		
Adverse event severity	**PPS (n=59), n (%)**	**Placebo (n=62), n (%)**
Grade 0 (none)	9 (15)	11 (18)
Grade 1 (mild)	8 (14)	7 (11)
Grade 2 (moderate)	30 (51)	34 (55)
Grade 3 (severe)	12 (20)	10 (16)
Parsons *et al.*, 1993[15] Adverse events occurring in more than one patient		
Adverse event	**PPS (n=74), n (%)**	**Placebo (n=74), n (%)**
Nausea	1 (1.4)	3 (4.1)
Diarrhoea	2 (2.7)	2 (2.7)
Vomiting	0	2 (2.7)
Sensation of euphoria	1 (1.4)	1 (1.4)
Watery eyes	1 (1.4)	1 (1.4)
Total reactions	12	19
Total patients (%)	7 (9)	10 (14)
Mulholland *et al.*, 1990[14]		
Adverse event	**PPS (n=54), n (%)**	**Placebo (n=56), n (%)**
Headache	1 (1.9)	2 (3.6)
Nausea	1 (1.9)	0
Indigestion	1 (1.9)	0
Increased perspiration	1 (1.9)	0
Severe mood swings	1 (1.9)	0
Suicidal ideation	1 (1/9)	0
Diarrhoea	0	2 (3.6)
Explosive diarrhoea	0	1 (1.8)
Severe joint pain	0	1 (1.8)
Skin rash (arms)	0	1 (1.8)
Itching	0	1 (1.8)
Total reactions	6	8
Total patients (%)	3 (6)	7 (13)
Parsons and Mulholland, 1987[16] after cross-over		
Adverse event	**PPS and PBO (n=62), n (%)**	
Skin rash	1 (16) unclear if PPS or PBO, or if before or after cross-over	

4.2.4 Quality assessment results for trials of PPS in IC/BPS

Table 11 presents the company's quality assessment of the four RCTs of PPS in IC/BPS.14-17

Details of the generation of the randomisation sequence, concealment of allocation, blinding, and imbalances in drop-outs were not reported in one of the published RCT reports (Parsons and Mulholland, 1987[16]) but were provided following communication with the trial author in Appendix Q of the CS.[1]

The ERG considers the company's quality assessment to be broadly accurate for three of the RCTs of PPS in IC/BPS.[14-16] However, the ERG considers some of the company's quality assessment judgements for the Sant *et al.* (2003) RCT[17] to be discrepant compared with the published report.[17]

Unlike the other three RCTs that report that the randomisation sequence was computer generated, Sant *et al.* (2003)[17] only report that a block randomisation by clinical site was performed, without details of the sequence randomisation generation method.

With respect to allocation concealment and blinding of participants and personnel, whilst the other three RCTs report that these aspects of trial design were undertaken, there is no record of allocation concealment being undertaken in the Sant *et al.* (2003) trial report and, although the Sant *et al.* (2003) trial is described as 'double-masked', unlike the other three RCTs, specific details of who was blinded is not reported.[17]

With respect to attrition bias, unlike the other three RCTs that report the number of drop-outs for PPS and placebo, Sant *et al.* (2003),[17] which used a factorial design resulting in four treatment groups, only reported the total number of drop-outs overall (20.6% across the four treatment groups – PPS, PBO, hydroxyzine, and PPS plus hydroxyzine). As such, it is unclear what attrition occurred in each of the treatment groups. Given the methodological quality issues in the Sant *et al.* (2003) RCT,[17] the ERG considers that the results from this trial should be interpreted with caution.

Table 11: Quality assessment of the pentosan polysulfate sodium RCTs relevant to the decision problem (adapted from Table 68 of the CS)

NICE criteria[23]	Sant *et al.*, 2003[17]	Parsons *et al.*, 1993[15]	Mulholland *et al.*, 1990[14]	Parsons and Mulholland, 1987[16]
An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups)	Yes (CS) Block randomised, so probably (ERG)	Yes Computer (ERG)	Yes Computer (ERG)	Yes Computer (ERG)
There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation)	Yes (CS) Not reported (ERG)	Yes	Yes	Yes
The groups were comparable at baseline, including all major confounding and prognostic factors	Yes	Yes	Yes	Yes
Based on your answers to the above, in your opinion was selection bias present?	No (CS) Unclear (ERG)	No	No	No
Likely direction of effect	NA	NA	NA	NA

NICE criteria[23]	**Sant *et al.*, 2003[17]**	**Parsons *et al.*, 1993[15]**	**Mulholland *et al.*, 1990[14]**	**Parsons and Mulholland, 1987[16]**
The comparison groups received the same care apart from the intervention(s) studied	Yes	Yes	Yes	Yes
Participants receiving care were kept 'blind' to treatment allocation	Yes	Yes	Yes	Yes
Individuals administering care were kept 'blind' to treatment allocation	Yes	Yes	Unclear	Yes
Based on your answers to the above, in your opinion was performance bias present?	No	No	No	No
Likely direction of effect	NA	NA	NA	NA
All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)	Yes	Yes	Yes	Yes
a. How many participants did not complete treatment in each group?	25 patients total (CS) Across PPS, hydroxyzine, PPS+ hydroxyzine, and PBO groups (ERG)	PPS: 9 Placebo: 9	PPS: 3 Placebo: 9	13 patients total
b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)	Yes	Yes	No: more patients in placebo group did not complete the trial	No: more patients in placebo group did not complete the trial
For how many participants in each group were no outcome data available?	Unclear	PPS: 6 Placebo: 4	Unclear	13 patients total
The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)	Yes	Yes	More dropouts in placebo group	More dropouts in placebo group
Based on your answers to the above, in your opinion was attrition bias present?	No	No	Unclear	No
Likely direction of effect	NA	NA	Unclear	NA

NICE criteria[23]	Sant *et al.*, 2003[17]	Parsons *et al.*, 1993[15]	Mulholland *et al.*, 1990[14]	Parsons and Mulholland, 1987[16]
The study had an appropriate length of follow-up	Yes	Yes	Yes	Yes
The study used a precise definition of outcome	Yes	Yes	Yes	Yes
A valid and reliable method was used to determine the outcome	Yes	Yes	Yes	Yes
Investigators were kept 'blind' to participants' exposure to the intervention	Yes	Yes	Yes	Yes
Investigators were kept 'blind' to other important confounding and prognostic factors	Yes	Yes	Yes	Yes
Based on your answers to the above, in your opinion was detection bias present?	No	No	No	No
Likely direction of effect	NA	NA	NA	NA

4.2.5 Summary of trials of PPS in the broader BPS population

The RCT by Holm-Bentzen *et al.* (1987)[18] was conducted in the UK and Denmark, whilst the RCT by Nickel *et al.* (2015)[19] was conducted in the USA and Canada.

Holm-Bentzen *et al.* (1987)[18] evaluated PPS 100 mg three times per day compared to PBO, whilst Nickel *et al.* (2015)[19] evaluated PPS 100 mg once per day or three times per day compared to PBO (three treatment groups).

The characteristics of the patients enrolled in both RCTs were broader than those indicated for PPS because the presence of Hunner's lesions and/or glomerulations were not part of the inclusion criteria.

In the RCT by Holm-Bentzen *et al.* (1987),[18] at four months there were no statistically significant between-group differences in symptoms, urodynamic parameters, cystoscopic appearance or mast cell counts (*p*-values or CIs, not reported). In the RCT by Nickel *et al.*, (2015),[19] at 24 weeks there was no statistically significant between-group difference in response defined as ≥30% reduction from the baseline in ICSI total score (*p*-values or CIs, not reported).

In summary, these two RCTs did not demonstrate evidence of a treatment effect of PPS in the broader BPS population.

In addition to the overall analysis, Nickel *et al.* (2015)[19] also reported a *post hoc* analysis of the primary end point in a subgroup of 94 participants who had objective findings of IC on cystoscopy meeting NIDDK criteria, done 30 days or more before enrolment or during the study. In this subgroup, the responder rate was greater with PBO than PPS (16/32, 50% vs. 10/29, 34.5%). These results were not presented in the CS.[1] In response to a request for clarification from the ERG, the company stated that the participants in Nickel *et al.* (2015)[19] were not stratified by NIDDK status at the randomisation of the trial, and breaking the randomisation in the *post hoc* analysis is likely to lead to bias in the estimates of relative treatment effect.[1]

The EMA also notes the limitations to both RCTs including patients in the broader BPS population.[8] The EMA also notes severe limitations of the Nickel *et al.*, (2015) study[19] (EPAR, page 93): "*patients with milder disease entering during a symptom flare, regression to the mean, introduction (inadvertent or not) of conservative therapy, which accentuated the benefits of placebo, and failure of clinical sites to keep patients in the trial are acknowledged by the CHMP. In addition, the results of Holm-Bentzen study are difficult to interpret as the GRA was not used as primary endpoint.*"

4.2.6 Included observational study on the increase in response rate over time

The CS[1] includes a section regarding the timing of response with PPS in IC/BPS, stating that "*Although some patients may experience improvements early in the PPS treatment process, others may not experience a clinical response until they have received 3–6 months of continuous PPS therapy*" (CS, page 81). In the CS, the company report on an increase in response rate over time reported in the single-arm study by Hanno (1997)[27] (Figure 4 below).

The ERG notes that of the 2809 participants recruited to the study by Hanno (1997),[27] 46% withdrew in the first three months and at 36 months there were only 149 (5%) participants left in the study. Therefore, the ERG considers that the results from this study should be interpreted with caution. In addition, the ERG considers the information difficult to interpret without a control group with which to estimate relative treatment effects.

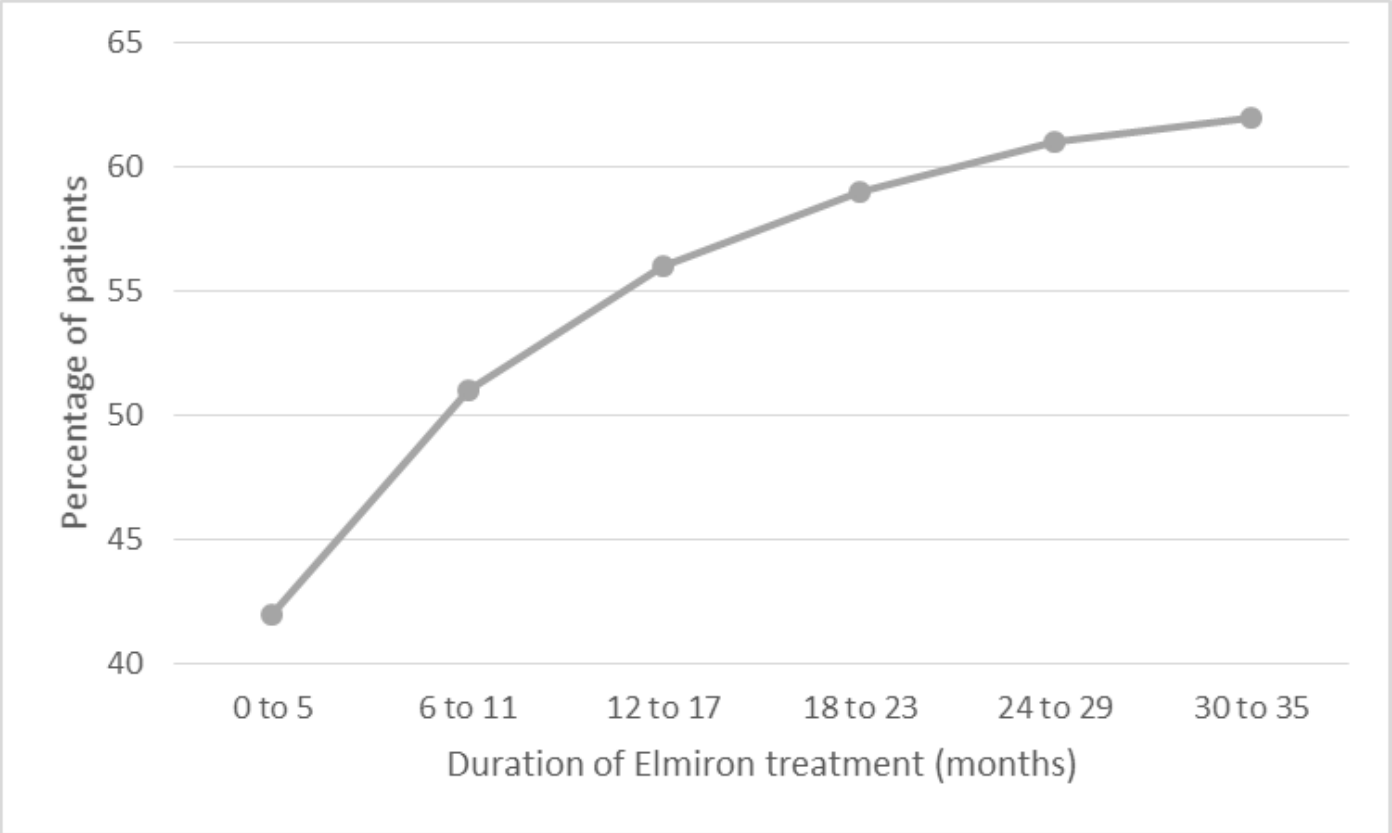


Figure 4: Percentage of patients with moderate or better improvement in patient global evaluation scale (reproduced from CS, figure 8, adapted by the company from Hanno 1997)

4.3 Critique of trials identified and included in the indirect treatment comparison

Details of the identification and methodology of the Uracyst® studies proposed to be included in an ITC analysis are described below. Details of the four PPS RCTs in IC/BPS, also included in the ITC are described in Section 4.2.

4.3.1 Search Strategy

The CS[1] (page 80) states that a systematic literature review (SLR) was conducted to identify studies to facilitate an ITC of PPS compared to other treatments included as comparators in the NICE scope.[2] Although not specifically stated in Section B.2.9 of the CS[1] (page 80), it appears that the trials proposed to be included in the ITC were identified from the SLR methods described in Section 4.1.

4.3.2 Study selection criteria

The CS states that the potential comparators in the review were defined more broadly than the NICE scope[2] to maximise the possibility of forming a network of trials. Although, Section B.2.9 of the CS[1] does not state explicitly whether the inclusion criteria for the ITC were the same as those for the clinical effectiveness review (CS, section B.2.1), it is stated that the inclusion criteria for the ITC were

the same in Appendix D.1 of the CS.[1] The ERG does not consider that any eligible trials have been missed.

4.3.3 Studies identified

The CS[1] (page 81) states that twelve trials met the inclusion criteria for the ITC. Of these, six trials compared PPS to placebo, and three trials compared Uracyst® to placebo instillation. The remaining three trials were excluded from the ITC as they did not include relevant comparators in order to construct a network. References to the excluded RCTs are not provided in this section; however, with reference to Table 66 in Appendix D of the CS[1] (page 163), it appears that the three studies identified as excluded are Tutolo *et al.*, (2017)[29]; Cervigni *et al.*, (2017)[30]; Gulpinar *et al.*, (2018).[31] The reasons for exclusion are presented in Table 12. The ERG considers the reasons for exclusion of these trials to be appropriate.

Table 12: List of studies excluded from the proposed ITC

Trial ID	Reason for exclusion
Tutolo *et al.*, (2017)[29]	Uracyst® compared to DMSO instillation
Cervigni *et al.*, (2017)[30]	iALuRil compared to DMSO instillation
Gulpinar *et al.*, (2018)[31]	Cystistat compared to Gepan

The PRISMA flow diagram reported in Appendix D of the CS[1] (page 162) shows that 15 RCTs (13 articles) were considered for inclusion in an ITC; this does not align with the information is reported on page 82 of the CS ("*Twelve trials met the inclusion criteria*"). In response to a request for clarification from the ERG, the company confirmed that this was an error and that 11 trials with 13 related citations had been considered for the ITC.

The CS presents proposed networks in Figures 9 and 10 on page 82 of the CS[1]. These figures list seven RCTs eligible for inclusion. These are five RCTs comparing PPS with placebo (Sant *et al.*, 2003;[17] Parsons *et al.*, 1993 ;[15] Parsons and Mulholland, 1987 ;[16] Nickel *et al.*, 2015 ;[19] and Holm-Bentsen *et al.*, 1987[18]), and two RCTs comparing Uracyst® with placebo (Nickel *et al.*, 2012[21] and Nickel *et al.*, 2010[20]). However, neither of these analyses were performed by the company due to considerable heterogeneity across the trials. For this reason, the company present only a meta-analyses of the data from the two Uracyst® trials versus placebo and compare this to the meta-analysis of data from the PPS versus placebo trials using the Bucher method.

Data from the PPS versus placebo trials have been critiqued in Section 4.1. In this section, we present a critique of the two Uracyst® versus placebo trials (Nickel *et al.*, 2012[21] and Nickel *et al.*, 2010[20]). These trials were selected for inclusion in order to compare PPS and Uracyst® in IC/BPS patients.

4.3.4 Quality assessment of studies included in the ITCs

It is unclear if the company performed quality assessment for Nickel *et al.* (2010)[20] and Nickel *et al.* (2012),[21] as neither the methods nor results of quality assessment were reported in the CS.[1] It is considered good systematic review practice for two reviewers either to independently perform quality assessment or to check assessed items; neither the quality assessment nor the checking was reported to have been done independently in the CS.[1] The ERG has completed the quality assessment for these two studies using the same criteria applied by the company for the main trials of interest which, although not referenced, is described as 'NICE criteria'. The ERG sought clarification with the company regarding this issue. The company's clarification response[1] stated that the items assessed were taken from the NICE Guidelines Manual.[23] These are appropriate criteria for assessing the methodological quality/risk of bias in RCTs.

As the CS does not present a quality assessment for Nickel *et al.* (2010)[20] and Nickel *et al.* (2012),[21] this was undertaken by the ERG using the quality assessment method applied by the company to the four RCTs of PPS in IC/BPS,[14-17] and is presented in Table 13 below.

Table 13: **Quality assessment of the trials used in the ITC**

NICE criteria	**Nickel *et al.*, 2010[20]**	**Nickel *et al.*, 2012[21]**
An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups)	Unclear - a predetermined randomization schedule	Yes - randomization schedule generated using a permuted block by a randomization statistician
There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation)	Unclear - this information is not provided.	Unclear - The women were randomized in a blinded fashion to the study treatment arms in a 1:1 ratio.
The groups were comparable at baseline, including all major confounding and prognostic factors	Yes	yes
Based on your answers to the above, in your opinion was selection bias present?	Unclear	Unclear
Likely direction of effect	Unclear	Unclear
The comparison groups received the same care apart from the intervention(s) studied	Yes	yes
Participants receiving care were kept 'blind' to treatment allocation	Yes	Yes
Individuals administering care were kept 'blind' to treatment allocation	Unclear - documentation of blinding does not specify, investigators, clinicians and participants	Unclear - documentation of blinding does not specify, investigators, clinicians and participants
Based on your answers to the above, in your opinion was performance bias present?	Unclear	Unclear
Likely direction of effect	Unclear	Unclear

NICE criteria	Nickel *et al.*, 2010[20]	Nickel *et al.*, 2012[21]
All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)	Yes	Yes
a. How many participants did not complete treatment in each group?	3 control, 4 intervention	9 control, 8 intervention
b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)	Yes	Yes
For how many participants in each group were no outcome data available?	1	0
The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)	Yes	Yes
Based on your answers to the above, in your opinion was attrition bias present?	No	No
Likely direction of effect	Unclear	Unclear
The study had an appropriate length of follow-up	Yes	Yes
The study used a precise definition of outcome	Yes	Yes
A valid and reliable method was used to determine the outcome	Yes	Yes
Investigators were kept 'blind' to participants' exposure to the intervention	Unclear	Unclear

NICE criteria	Nickel *et al.*, 2010[20]	Nickel *et al.*, 2012[21]
Investigators were kept 'blind' to other important confounding and prognostic factors	Unclear	Unclear
Based on your answers to the above, in your opinion was detection bias present?	Unclear	Unclear
Likely direction of effect	Unclear	Unclear

Details of the generation of random sequence was not reported by Nickel *et al.* (2010)[20], but in Nickel *et al.* (2012)[21] it was reported that the randomisation schedule was generated using a permuted block. In Nickel *et al.* (2010)[20] it was not clearly stated that concealment of allocation had taken place, although in Nickel *et al.* (2012)[21] it was stated that the women were randomised in a blinded fashion, implying that concealment of allocation had taken place. Although both studies were described as double-blind, and methods for blinding regarding active or vehicle controlled instillation was described in Nickel *et al.* (2010),[20] neither clearly outlined which study personnel were blinded to the study arms. With respect to attrition bias, the number of drop-outs in each arm are provided, and appear balanced in both studies. On the basis of the quality assessment, the ERG concludes that these trials were of moderate to low quality.

4.3.5 Critique of studies included

4.3.5.1 Study designs

Both trials report that they were multicentre, double-blind, and randomised. Both trials appear to have been conducted in Canada, and the number of centres was reported as 12 in Nickel *et al.* (2010)[20], but was not reported in Nickel *et al.* (2012)[21]. Nickel *et al.* (2010)[20] is described as an inactive vehicle-controlled study, parallel group pilot evaluation, whilst Nickel *et al.* (2012)[21] is described as an inactive control trial, parallel group evaluation. Detail regarding when the trials were initiated and completed are not available in the CS or the trial papers. The studies were conducted relatively recently and broadly represent best practice in the UK.

4.3.5.2 Population characteristics

Eligibility criteria of the included studies were not outlined in the CS. Nickel *et al.* (2010)[20] specified that patients had to be 18 years old or over, but this was not specified in Nickel *et al.* (2012).[21] There did not appear to be an age cut off for either trial. Nickel *et al.* (2012)[21] only included women, whereas Nickel *et al.* (2010)[20] included both men and women, although only one male was

randomised into the study and was part of the control group. Both trials were described as being in the IC/PBS population; however, the diagnostic criteria did not include the presence of ulcers or petechial haemorrhage on cystoscopy (see Table 14). The CS[1] (Table 30) defined the populations in both trials as BPS.

Table 14: **Diagnostic eligibility criteria for the included studies derived from study reports**

Nickel *et al.* (2010)[20]	**Nickel *et al.* (2012)[21]**
Clinical diagnosis of IC/PBS The diagnosis of IC/PBS was consistent with current clinical definitions, including the diagnostic criteria described in the IC Data Base Study,[12] as well as the most recent definition of IC/PBS described at the NIH Urologic Chronic Pelvic Pain consensus (Baltimore, December 2007). IC/PBS was diagnosed on the basis of pelvic pain, pressure, or discomfort perceived to be related to the urinary bladder accompanied by at least one other urinary symptom, such as urgency or frequency.	Diagnosed or re-diagnosed with IC/BPS within the previous 2 years; had a subject-reported average urinary frequency of 8 times/24 hours during the screening period, as captured by a 3-day diary; had a pain/pressure/discomfort score of 40-80 mm on a pain visual analogue scale (VAS).

Baseline characteristics appeared to be broadly comparable across the trial arms in both trials. Details of the ethnicity of the patients were not reported in the CS or in the trial report for Nickel *et al.* (2010)[20], whilst the ethnicity of the patients was reported in Nickel *et al.* (2012)[21] and was comparable across trial arms, and appeared to be broadly generalisable to the UK population.

The eligibility criteria detailed in the trial papers included a diagnosis of IC/BPS, but do not report that Hunner's lesions and/or glomerulations are part of the diagnosis. However, patients in these trials are defined in the CS [1] as patients with bladder pain syndrome with Hunner's lesions and/or glomerulations. Therefore, it is not clear that the patients in either trial met the criteria for this NICE scope (see Table 14). The ERG also notes that neither the CS[1] or the individual trial papers report numbers of patients overall or in each arm with either Hunner's lesions or glomerulations. (see Section 4.2.1 for further discussion).

4.3.5.3 Intervention characteristics

The intervention characteristics for the RCTs are listed in Table 15. The intervention appears to be consistent with the NICE scope[2] in terms of dosing and administration, and is broadly comparable with UK practice.

Table 15: Characteristics and results of Uracyst® trials (adapted from CS, Table 30 p.85-86)

Study	**Nickel *et al.* (2010) [20]**		**Nickel *et al.* (2012) [21]**	
Study type	Prospective, randomised, double-blind, inactive vehicle-controlled study		Multicentre, double-blind, inactive control trial	
Population	Adult patients with BPS		Women with BPS	
Intervention	2% sodium chondroitin sulphate (Uracyst®)	Intravesical vehicle control	2% sodium chondroitin sulphate (Uracyst®)	Inactive control instillation
Sample size	33	32	50	48
Mean follow-up time	6 weeks (12 week study with 6 weeks of treatment and 6 weeks of follow-up)		11 weeks	

Abbreviations: GRA, global response assessment; ICPI, interstitial cystitis problem index; ICSI, interstitial cystitis symptom index; NR, not reported; SD, standard deviation.

4.3.5.4 Outcome assessment

The CS reports some outcome data for the trials (see Table 16). These are consistent with those outlined in the NICE scope.[2] The CS does not report information about the methods for assessing outcomes in the trials. In Nickel *et al.* (2010),[20] patients underwent a six week treatment period, followed by a 6 week follow up period. The primary outcome was the number of patients in each group who moderately or markedly improved on the Global Response Assessment (GRA) scale. Outcomes were reported at weeks 7 and 12, with 7-week outcomes as the primary measure. Secondary efficacy endpoints were the O'Leary-Sant interstitial cystitis Symptom Index/Problem Index (ICSI/ICPI), the Female Sexual Function Index (FSFI), the Short Form 12 quality of life Questionnaire (SF-12), daily urinary frequency, the Likert pain scale, and safety outcomes. In Nickel *et al.* (2012)[21] there was a 7 week treatment period, followed by a 4 week follow up period, with primary and secondary endpoints assessed at week 11. The primary outcome was GRA; the secondary outcomes were the ICPI, average daily urinary frequency, average urine volume per void, average daily urgency episodes and pain VAS score.

Although the timing of the primary outcome differed in the trials, comparable data for end of follow up was available from the trial reports, and these data were used in the meta-analyses. Only data on the primary outcome, GRA, and secondary outcomes ICSI and ICPI were reported and meta-analysed in the CS. The definitions of the outcomes and follow up time appear comparable across the trials, although details of the outcome assessor are not available in the trials. None of the findings were statistically significant (for p-values see Table 16, CIs not reported).

An inconsistency between the data reported in the CS and those reported in the original trial reports was noted by the ERG. The mean (SD) ICSI at baseline for Nickel *et al.* (2010)[20] for the intervention group was reported as 12.4 (3.26) in the CS; however, these data were reported as 13.8 (3.55) in the trial paper.

Table 16: Results of Uracyst® trials (adapted from CS, Table 30 p.85-86)

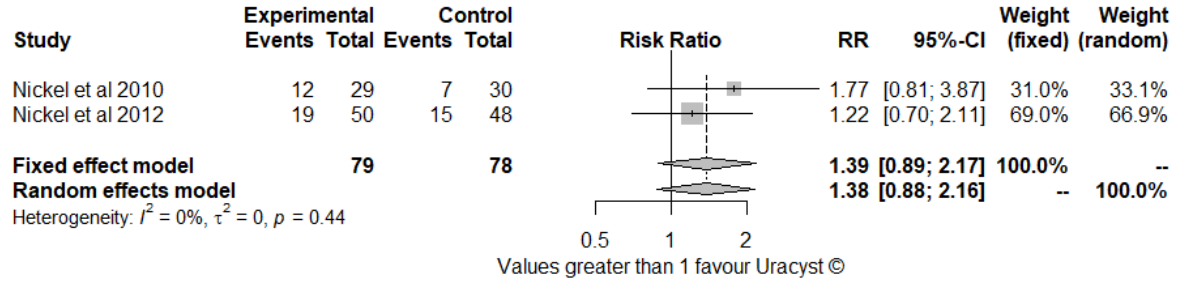
Study	Nickel *et al.* (2010) [20]		Nickel *et al.* (2012) [21]	
GRA at follow-up (SD)	Week 7: 13 (39.4) Week 12: 12 (41.4)	Week 7: 7 (22.6) Week 12: 7 (23.3)	NR	
GRA response at follow-up (%)	NR		Yes: 19 (38.0) No: 31 (62.0)	Yes: 15 (31.3) No: 33 (68.8)
P value (between groups)	Week 7: 0.1470 Week 12: 0.1381		0.4828	
Mean ICSI at baseline (SD)	12.4 (3.26)	14.7 (3.02)	12.9 (3.40)	12.8 (3.46)
Mean ICSI change from baseline (SD)	Week 7: -2.8 (3.68) Week 12: -2.7 (4.07)	Week 7: -2.8 (2.39) Week 12: -3.2 (3.5)	NR	
Mean ICSI at follow-up (SD)	NR		9.7 (4.99)	9.7 (4.92)
P value (between groups)	Week 7: 0.8458 Week 12: 0.7069		0.9536	
Mean ICPI at baseline (SD)	12.4 (3.26)	12.9 (2.28)	12.4 (2.69)	11.7 (3.00)
Mean ICPI change from baseline (SD)	Week 7: -2.9 (3.26) Week 12:-3.0 (3.75)	Week 7: -3.1 (3.23) Week 12: -2.9 (3.63)	NR	
Mean ICPI at follow-up (SD)	NR		7.9 (4.59)	8.3 (4.51)
P value (between groups)	Week 7: 0.7668 Week 12: 0.8771		0.4656	

Abbreviations: GRA, global response assessment; ICPI, interstitial cystitis problem index; ICSI, interstitial cystitis symptom index; NR, not reported; SD, standard deviation.

Pairwise meta-analysis of effectiveness

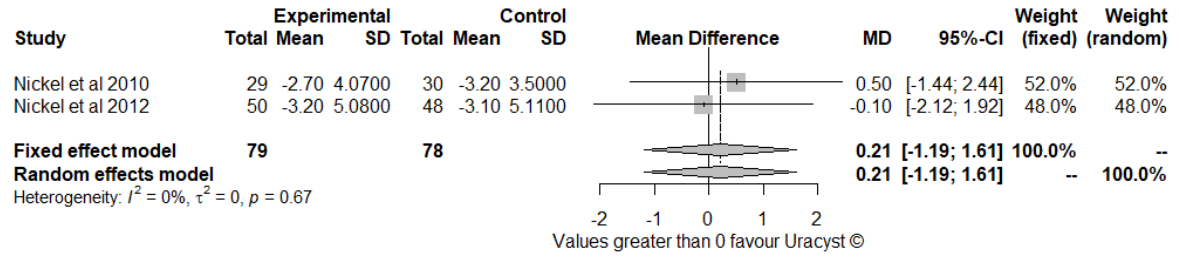
The company presented pairwise meta-analyses across the two Uracyst® RCTs[20, 21] for GRA, ICSI and ICPI (CS, figures 13 to 15). The forest plots for these analyses are presented in Figure 5, Figure 6 and Figure 7 below. For GRA, the fixed effect RR of 1.39 (95%CI: 0.88 to 21.7) was applied by the company in the economic model.

The pooled estimate was used to compare to the pooled GRA estimate from the pairwise meta-analyses across four RTCs of PPS in IC/BPS[14-17] based on the Bucher method. Further details of this are presented in Section 4.4 of this ERG report.



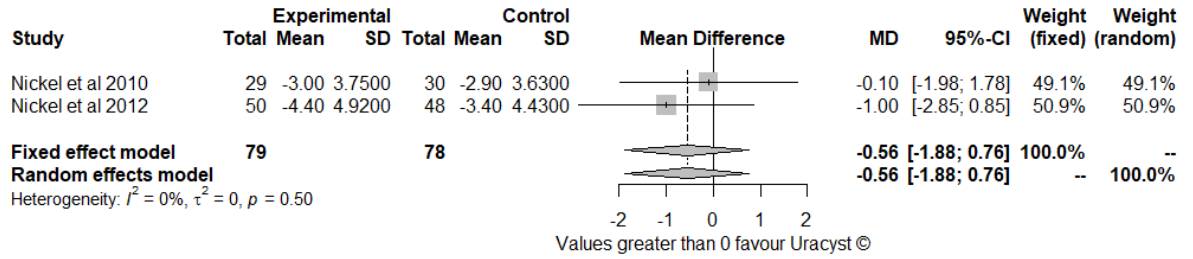
Reproduced from the CS page 86.1

Figure 5: Meta-analysis of Global Response Assessment in RCTs of Uracyst® in /BPS (risk ratio) (CS, figure 13)



Reproduced from the CS page 87.[1]

Figure 6: Meta-analysis of for mean change in Interstitial Cystitis Symptom Index in RCTs of Uracyst® in /BPS (risk ratio) (CS, figure 14)



Reproduced from the CS page 87.[1]

Figure 7: Meta-analysis of for mean change in Interstitial Cystitis Problem Index in RCTs of Uracyst® in /BPS (risk ratio) (CS, figure 15)

Adverse events

Adverse events for the Uracyst® trials were not reported in the CS. Nickel *et al.* (2010)[20] reported that 76.9% of the patients in the study reported at least 1 adverse event (AE) , 87.5% (28/32) of the control group reported 86 AEs compared with 66.7% (22/33) of the treatment group reported 67 AEs. Most AEs were reported as being mild in severity (56 and 45 for control and treatment groups respectively). There were 25 and 17 moderate AEs reported for the control and treatment groups, respectively, whereas 10 severe AEs were reported overall (5 in each group). Only nine intervention-related AEs were reported in three patients of the control group, compared with two intervention-related AEs in one patient in the treatment group. Intervention-related AEs were considered mild in the treatment group and mostly mild or moderate in the control group.

Nickel *et al.* (2012)[21] reported that 70.4% of the patients in the study experienced one or more AE (57.1% reported as mild intensity). However, the investigators reported that there was no "difference" was reported between the control group (71.4%) and the treatment group (69.4%). Only 7.6% of the AEs were intervention-related (10.3% in the control group and 5.2% in the treatment group). Four unrelated but serious, AEs occurred in 3 patients (suicide ideation and angina in 2 patients in the control group; and rectal bleeding and chronic colitis in 1 patient in the active treatment group). One patient in the active treatment group discontinued because of an unrelated AE. *P*-values or CIs were not reported.

4.4 Critique of the pairwise meta-analyses and indirect treatment comparison

Results of the pairwise meta-analyses undertaken by the company are presented in Sections 4.2.2 and 4.3.3 of this ERG report.

The ERG has some concerns with the company's meta-analyses (CS, Section B.2.8.1).[1] The aim of the meta-analysis was to determine the efficacy of PPS for the treatment of IC/BPS in comparison

with placebo. In general, the aim of a meta-analysis is to generate an estimate of the treatment effect on an additive scale that can be transported and used to estimate absolute risk in a target population. In addition, an objective in this submission is to generate a (posterior) distribution for the treatment effect and an estimate of the baseline effect that can be used together to represent uncertainty about absolute responses to treatment in the economic model:

- Table 28 of the CS[1] presents a meta-analysis of four studies on the risk difference scale for GRA. While a meta-analysis of risk difference may be appropriate when the baseline event rates are similar among the studies, treatment effects are more likely to be additive on a relative scale such as the log-odds ratio or log-relative risk.
- The company concluded that "There was a high degree of homogeneity in this sensitivity meta-analysis …" based on Cochran's Q value. The ERG has concern with the use of Cochran's Q value to assess and conclude homogeneity of relative treatment effects across studies,[32] and has a preference for estimating the between-study standard deviation and its uncertainty. In addition, it is unclear to the ERG why the company refers to this meta-analysis as a sensitivity analysis.
- The company's misinterpretation of Cochran's Q value is repeated when they include two additional studies with a broader BPS population[18, 19] that the ERG recognises do not satisfy the inclusion/exclusion criteria for the assessment. The company claims that "there is no indication of heterogeneity" rather than the more appropriate interpretation that there is insufficient evidence to reject the null hypothesis that there is homogeneity of treatment effects. Furthermore, and somewhat contradictory, the company goes on to state that the results were heterogeneous.
- It is not clear whether the meta-analysis presented in Table 29 of the CS[1] is based on a fixed or random effects model, and the predictive distribution of an effect is not provided. The use of a fixed effect meta-analysis is appropriate if interest is in a conditional inference of whether treatment had an effect in the available studies or if all of the factors that could affect the effect size on an appropriate additive scale are the same in all study populations. When there is reason to believe that the effect size may not be identical in the available or any future studies that might be conducted then a random effects meta-analysis should be performed; the choice between a fixed effect and random effects model should not be based on a test of heterogeneity of treatment effects.
- The company has used standard frequentist methods assuming asymptotic normality which may not be optimal given the samples sizes used and the number of observed events in the available studies. An exact analysis of the data using a Binomial likelihood and generation of the (posterior) distribution for the treatment effect could have been done using Bayesian methods.

- It is unclear what the relevance is of the lower limit of the 95% confidence interval for the pooled estimate of the absolute difference in GRA response being less than 5%. The lower limit of the 95% confidence interval suggests that it is unlikely that treatment effects smaller than 6.4% (not 6.3% as stated in the CS) are consistent with the data. The ERG notes that this does not mean that the 95% confidence interval contains clinically meaningful values which would require specification of and comparison with a minimum clinically important effect size.

Section B.2.8.2 of the CS[1] describes a meta-analysis conducted by Hwang *et al.* (1997).[33] The CS erroneously states that the meta-analysis included Sant *et al.* (2003),[17] which is impossible given that Sant *et al.* (2003)[17] was published after Hwang *et al.* (1997).[33] The CS[1] reports arm-based pooled estimates of response rates which is generally not recommended and implied wrongly that Hwang *et al.* (1997)[33] used these to make inferences about treatment effects.

The CS[1] discusses the impact on outcomes as a consequence of patients being enrolled in a clinical trial. The ERG considers the placebo effect to be irrelevant in the context of estimating a relative treatment effect that is transportable assuming that the relative treatment effect is estimated on an appropriate additive scale. On the basis that the relative treatment effect is estimated on an appropriate additive scale then it is necessary only to specify the absolute effect on the same scale for the control treatment when used in clinical practice in the target population in order to generate absolute effects.

The company presents data from Hanno (1997),[4] describing the percentage of patients with moderate or better improvement in patient global evaluation scale following treatment with PPS at six-monthly intervals over three years. The ERG considers the information difficult to interpret without a control group with which to estimate relative treatment effects.

Overall, the ERG accepts the company's argument that an unbiased comparison between PPS capsules and all relevant comparators was not possible using a conventional network meta-analysis because the studies of Uracyst® included patients in the wider BPS population, the placebo in the Uracyst® studies was a placebo instillation whereas the comparator in the PPS trials was a placebo capsule, and the timing of assessments differed between studies.

Nevertheless, in order to satisfy the NICE scope, the company provided an ITC between PPS and Uracyst® linked by the placebos using the Bucher method.[34] The ERG has a preference for performing a simultaneous comparison between treatments using a Bayesian network meta-analysis for the following primary reasons: (1) the Bucher approach allows for separate and unrelated meta-analyses for the effect of PPS versus placebo and the effect of Uracyst® versus placebo whereas a single model incorporates a common random effect, (2) the posterior distribution for the effect of PPS

versus Uracyst® will not follow any standard parametric distribution whereas the Bucher approach involves an assumption of asymptotic normality when making inferences, and (3) the relative treatment effects of PPS versus placebo and Uracyst® versus placebo will be correlated and this will induce correlation between absolute responses to treatment when combined with an external estimate of the baseline response. In addition, in the absence of evidence that there is no heterogeneity of treatment effects between studies, the ERG has a preference for a random effects model allowing for uncertainty in the estimate of the between-study standard deviation and estimation of the predictive distribution of treatment effect which is straightforward and exact using a Bayesian approach.

The company summarised the GRA data on the relative risk scale to characterise uncertainty about relative treatment effects for use in the economic model. The ERG notes that treatment effects should be estimated on an additive scale and that if treatment effects are additive on one scale such as the absolute scale as presented in Tables 28 and 29 of the CS then they cannot be additive on another scale as presented in Figure 11 of the CS. Estimation of the treatment effect may be appropriate on the risk difference scale if the GRA response rate is assumed to be zero in clinical practice or if the baseline event rates are similar among the studies being analysed; otherwise, the ERG has a preference for analysing the data on the logit scale.

The CS presents results from fixed effect and random effects models in order to estimate the relative effects of PPS versus placebo and of Uracyst® versus placebo using a frequentist approach. A frequentist approach assumes that the variance of the pooled estimate is known and ignores uncertainty in both the within-study estimate of variance and the between-study estimate of variance. Accurate inferences require reasonably large studies with which to estimate the within-study variance precisely and a reasonably large number of studies (i.e., at least five) with which to estimate the between-study variance. Consequently, the ERG suggests that there is insufficient information with which to assess heterogeneity as claimed by the company and that the results of the random effects models presented in Figures 11-15 of the CS should be treated with caution. Nevertheless, the ERG has a preference for random effects models except when making conditional inferences or when it is known that studies are estimating the same underlying treatment effect. This could be done using a Bayesian approach incorporating external information about the between-study standard deviation.

4.5 Additional work on clinical effectiveness undertaken by the ERG

No additional work was required to be undertaken by the ERG.

4.6 Conclusions of the clinical effectiveness section

The ERG considers that the company's search strategy is sufficiently comprehensive to retrieve important citations relating to clinical effectiveness and safety of pentosan polysulfate (PPS) for

treating Interstitial cystitis/bladder pain syndrome (patients with bladder pain syndrome with Hunner's lesions and/or glomerulations) (IC/BPS).

The four RCTs of PPS in IC/BPS were relevant to the decision problem outlined in the final NICE scope. Three of the RCTs of PPS in IC/BPS (Parsons *et al.*, 1993, Mulholland *et al.*, 1990 and Parsons and Mulholland, 1987) were considered by the ERG to be of good methodological quality. However, the ERG considered one RCT (Sant *et al.*, 2003) to be unclear regarding: allocation concealment, details of who was blinded, and numbers of patients withdrawing from treatment groups. As such, that the results from this trial should be interpreted with caution.

The ERG notes potential issues surrounding study power and sample size as three of the RCTs of PPS in IC/BPS did not prospectively define the sample size (Parsons *et al.*, 1993; Mulholland *et al.*, 1990; Parsons and Mulholland, 1987), and the one RCT which reported a power calculation failed to recruit the target number of patients (Sant *et al.*, 2003).

The ERG also notes limitations in the reporting of outcome data in the PPS RCTs trial reports. Interval estimates (CIs) were not reported and, where between-group differences were reported as not statistically significant, *p*-values were often not reported.

All four RCTs of PPS in IC/BPS were multicentre trials conducted in the USA and published between 1987 and 2003. The ERG notes that there is some author commonality across all four RCTs of PPS in IC/BPS and that subsequently, there has not been any further published study undertaken by an independent study group which has attempted to validate the results of the four RCTs of PPS in IC/BPS.

The between-group difference in the proportions of patients with a patient-reported >50% improvement in global response assessment (GRA) at three months was reported as being statistically significant in favour of PPS by two RCTs (Parsons et al, 1993, PBO 16% vs PPS 32%, $p=0.01$; and Mulholland *et al.*, 1990, PBO 13% vs PPS 28%, $p=0.04$; CIs not reported), but the between-group difference in the proportions of patients with a GRA score of six to seven at three months was reported as not statistically significant by one RCT (Sant *et al.*, 2003, PBO 18% vs. PPS 34%, $p=0.064$). As GRA was not assessed in one RCT (Parsons and Mulholland, 1987), the proportions of patients with a >50% improvement in non-VAS pain was used as a proxy for GRA in the analysis undertaken by the company. The between-group difference in non-VAS pain reported by Parsons and Mulholland (1987) was statistically significant (PPS 44% vs. PBO 15%, $p=0.02$; CI not reported).

The between-group difference in in the proportions of patients with a >50% improvement in non-VAS pain at three months was reported as being statistically significant in one other RCT (Parsons *et al.*, 1993; $p=0.005$). However, the between-group the between-group difference in mean non-VAS pain scores were reported as not being statistically significant at three months for two other RCTs (Sant *et al*., 2003; Mulholland *et al.*, 1990; *p*-values or CIs, not reported).

The between-group difference in the O'Leary-Sant Interstitial Cystitis Symptom Index and Problem Index mean scores at three months were reported as being not statistically significant in one RCT (Sant *et al.*, 2003, *p*-values not reported). There were no statistically significant between-group differences in mean daily urinary frequency at three months reported by two RCTs (Sant *et al.*, 2003, *p*-value not reported; Parsons and Mulholland, 1987, $p=0.06$). There were no statistically significant between-group differences in mean urinary volume and void outcomes at three months reported by two RCTs (Parsons *et al.*, 1993; Mulholland *et al.*, 1990; *p*-values not reported), and one RCT did not report whether the between-group difference was significant or not, or a *p*-value for the between-group difference (Parsons and Mulholland, 1987). There were no statistically significant between-group differences in mean nocturia at three months reported by two RCTs (Mulholland *et al.*, 1990, *p*-value not reported; Parsons *et al.*, 1993, no data reported), and one RCT did not report whether the between-group difference was significant or not, or a *p*-value or CI, for the between-group difference (Parsons and Mulholland, 1987).

Safety data for PPS were presented in the CS from each of the individual RCTs of PPS in IC/BPS, and the company concluded that PPS is well tolerated. Common adverse events in the SmPC are: headache, dizziness, nausea, diarrhoea, dyspepsia, abdominal pain, abdominal enlargement, rectal haemorrhage, peripheral oedema, alopecia, back pain, asthenia, and pelvic pain. However, clinical advice received by the ERG based on named patient use is that AEs are rare with PPS.

The ERG has some concerns with the pairwise meta-analyses that were performed by the company and reported in the CS (the choice of scale for the analysis, the use of hypothesis testing to assess heterogeneity, and the use of a fixed effect model in the absence of evidence that there is not between study heterogeneity). The ERG accepts the arguments suggested by the company for not performing an NMA. Nevertheless, an ITC between PPS and Uracyst® was required and the company did this using the Bucher method, with the placebos as the reference treatment. While neither an NMA nor the Bucher approach are ideal in this case, the ERG does not believe that the Bucher approach mitigates all of the concerns associated with performing an NMA, including: not using a single model to incorporate random effects; making the assumption of asymptotic normality when making inferences and characterising uncertainty about the relative treatment effect used in the economic model.

5 COST EFFECTIVENESS

5.1 ERG's comment on company's review of cost-effectiveness evidence

This section presents a review of the cost-effectiveness evidence reported in the CS[1] for pentosan polysulfate sodium (PPS) for treating IC/BPS (defined as patients with bladder pain syndrome with Hunner's lesions and/or glomerulations).

5.1.1 Objective of cost effectiveness review

The company undertook a systematic literature review in order to identify cost-effectiveness evidence for IC/BPS and BPS treatments.

Two searches were performed to identify economic evaluations of IC/BPS and BPS. The following databases were searched for economic evaluations in June 2018: MEDLINE [via Ovid], MEDLINE In-Process & Other Non-Indexed Citations [via Ovid], Embase [via Wiley], NHS EED [via Wiley]. The company carried out supplementary searches within health technology appraisals via the NICE website.

In the company's clarification response (question B1), the company reported that publication date limits were not applied to the economic and cost-effectiveness evaluations searches. The NHS EED database coverage is limited to 1995-2014 whereas limits of 1992-1994 and 2015-present were applied in the MEDLINE and Embase search. The reasons and implications of not including all years in MEDLINE and Embase were not given. The ERG is unable to confirm if any key economic evaluations have been missed as a result applying these limits.

The company performed two searches to identify health-related quality-of-life studies for IC/BPS and BPS. Details of these searches were provided in response to a request for clarification from the ERG (question B3).[35] The following three sources were searched in June 2018: MEDLINE [via Ovid], MEDLINE In-Process & Other Non-Indexed Citations [via Ovid] and Embase [via Ovid]. The company cross-checked lists of included articles with records from the electronic searches. The ERG considers that the searches are sufficiently comprehensive to retrieve all the eligible studies.

The company performed two searches to identify cost and resource use evidence for IC/BPS and BPS. Details of these searches were provided in response to a request for clarification from the ERG (question B3).[35] The following two sources were searched in June 2018: MEDLINE [via Ovid], MEDLINE In-Process & Other Non-Indexed Citations [via Ovid] and Embase [via Ovid The company cross-checked lists of included articles with records from the electronic searches. The ERG identified one study that should have been included.[36] The ERG cross-checked the study against the

MEDLINE and Embase search results and confirmed that the record would have been missed by the company's searches.

5.1.2 The inclusion and exclusion criteria used in the study selection

The inclusion and exclusion criteria for the cost-effectiveness review was included in Appendix G (Table 69) of the CS[1]. The ERG believes that the company's criteria were acceptable in order to identify relevant studies of cost-effectiveness in the population of interest. The ERG believes that the exclusion criteria used for cost and healthcare resource use studies could have excluded potential studies that could provide data for model inputs for costs and resource use. The ERG believes that the exclusion criteria (no intervention/comparator) used for HRQoL studies could have excluded potential studies reporting baseline quality of life data in the population defined in the NICE scope.

5.1.3 Findings of the cost effectiveness review

Four studies were identified for full text screening with only one study identified and included within the cost-effectiveness review. This study, conducted by Cervigni (2017)[30],was a within trial economic evaluation of iAluRil® vs. DMSO and provided baseline EQ-5D values for patients with IC/BPS which were based on Italian population values. Three studies were excluded at the full text stage as they were not economic analyses. The details of these three excluded studies were provided by the company during the clarification process (question B41). No studies assessing the cost-effectiveness of PPS for the treatment of IC/BPS were identified. Additional searches undertaken to identify cost and healthcare resource use studies identified two studies for data extraction, one of which was the aforementioned Cervigni 2017[30]. The ERG notes that the CS[1] and the clarification responses[1] state that the three studies excluded from the main cost-effectiveness review were excluded as they only contained costs and healthcare resource use data. The ERG is unsure why these studies were therefore not identified and included in the additional costs and healthcare resource use reviews. In addition, an ad hoc search conducted by the ERG identified a costing study related to treatment costs of IC/BPS in Austria[36] which meets the company's inclusion criteria for cost and healthcare resource use studies, yet was not included. The ERG is unsure why a study with no intervention /comparator[37] was included in the HRQoL studies when the exclusion criteria states that these studies would be excluded.

5.1.4 Conclusions of the cost effectiveness review

The ERG is satisfied that the identified published cost-effectiveness study[3] is not appropriate to address the decision problem in the NICE scope,[2] and therefore that the development of a de novo model is appropriate. However, the ERG has some concerns about the date limits applied to the company's economic searches and is unable to confirm if any key economic evaluations have been missed as a result of applying these limits. In addition, the ERG has some concerns with the quality of

the searches undertaken for additional studies of cost and healthcare resource use data and HRQoL studies.

5.2 Summary and critique of company's submitted economic evaluation by the ERG

Please note that the company submitted a revised model following the clarification request and it is this model that is referred to throughout the report unless otherwise specified. The revisions made in this model were mainly corrections of errors in the implementation of the model and did not concern the model structure, assumptions or data sources, with the exception of the life-table data being updated to the most recent dataset available.

5.2.1 NICE reference case checklist

Table 17: Compliance with the NICE reference case[38]

Element	Reference case	ERG comments
Defining the decision problem	The scope developed by NICE	The population modelled is adults with BPS characterised by either glomerulations or Hunner's lesions, which is consistent with the NICE scope and the licensed indication for PPS. The evidence on the effectiveness of PPS compared to BSC is taken from trials in the relevant population.[14-17] However, the evidence on the effectiveness of BI compared to placebo, comes from the broader population of patients with BPS which is not restricted to those with either glomerulations or Hunner's lesions. The model evaluates the cost-effectiveness of PPS separately in the subgroup able to receive BIs and the subgroup who are contraindicated or unable to tolerate BIs as both the comparators and subsequent treatments differ in these populations. The scope explicitly states that the economic modelling should include the costs associated with diagnostic testing for glomerulations or Hunner's lesions in people with bladder pain syndrome who would not otherwise have been

		tested. The model does not incorporate any costs for diagnostic testing. However, the ERG is satisfied that this is reasonable based on the advice provided by their clinical experts which stated that the relevant test would be carried out as part of the standard diagnostic pathway, whether or not PPS was being considered as a treatment option.
Comparator(s)	As listed in the scope developed by NICE	• BIs in the population able to receive BIs • BSC in the population unable to receive BIs These are consistent with the scope
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	The model estimates direct health effects for patients but not carers which is considered by the ERG to be reasonable in this case. In estimating the QALYs, the model does not capture AEs of either PPS or BIs; however, this is not considered to have significantly biased the assessment of cost-effectiveness
Perspective on costs	NHS and PSS	The model includes only NHS costs as the CS[1] states that PSS costs are not relevant. This is considered to be a reasonable deviation from the reference case in this case.
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The submitted model provides a cost-utility analysis with outcomes presented as the incremental cost per QALY gained for two comparisons; • PPS versus BIs in the population able to receive BIs • PPS versus BSC in the population unable to receive BIs The ERG considers this approach to be reasonable given that BIs are the current standard of care and BSC would only be given to those patients who are unable to receive BIs.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The base-case analysis uses a 20-year time-horizon; sensitivity analyses are provided using a lifetime horizon. The ERG considers the lifetime horizon more appropriate for the reference case analysis given that the survival function used to extrapolate time to discontinuation predicts that 18% of patients remain on treatment at 20 years.
Synthesis of	Based on systematic	The estimates of treatment effect for PPS versus BSC are

evidence on health effects	review	based on a systematic review and meta-analysis of RCTs. The estimate of treatment effect for BIs versus BSC are based on a systematic review and meta-analysis of RCTs for commercial BIs. Evidence was only identified for one form of BIs. The model assumes that all BIs are equally efficacious. A simple indirect comparison, using the Bucher method[34], has been used to compare PPS to BSC rather than a network meta-analysis.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults.	Health effects are expressed in QALYs with utility values based on the EQ-5D.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	EQ-5D-5L responses were measured in a patient survey along with a measure of disease severity (ICSI).[39] Utility scores for the model were estimated by mapping from ICSI scores to EQ-5D.[39] The measure of efficacy in the model is the response rate. The expected ICSI scores for responders and non-responders were estimated from a single trial arm (PPS arm of Sant 2003[17]) and were applied universally to all patients in the model according to their response to treatment.
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	The EQ-5D-5L responses from the survey[39] were cross-walked to the UK EQ-5D-3L valuation set using the mapping function developed by van Hout.[40] This is consistent with the approach recommended for reference case analyses according to NICE's current position statement on this topic.[41]
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health	No weighting of QALYs has been applied. This is consistent with the NICE reference case.

	benefit	
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	The CS[1] states that resource use has been costed using standard prices relevant to the NHS such as NHS reference costs and list prices. The price of PPS was sourced from the manufacturer. The price of one of the BIs was also sourced from the manufacturer, but all other drug prices were based on public list prices.
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	In line with the NICE reference case, costs and health effects are discounted at a rate of 3.5% per annum.

5.2.2 Model structure

The company's model is a discrete event simulation (DES) which estimates the mean costs and QALYs for a cohort of patients by simulating outcomes for 10,000 individuals with identical characteristics at baseline. The model simulates the clinical events occurring over the time horizon for each individual and uses these to predict lifetime costs and QALYs for each individual. Even though patients have identical characteristics at baseline, their path through the model is allowed to vary stochastically (i.e. according to chance) by sampling the time that the various possible events will occur for each individual from time-to-event distributions. The costs and QALYs expected for an average patient are estimated by taking the average across the simulated cohort of individuals. This provides a stable estimate of the expected costs and QALYs if outcomes for a sufficient number of individuals are sampled. The outcomes are simulated for the whole cohort for each treatment option (i.e. once for PPS, BI, and BSC) and then the average outcomes for each treatment option are compared to provide estimates of the incremental costs and QALYs between alternative treatment options.

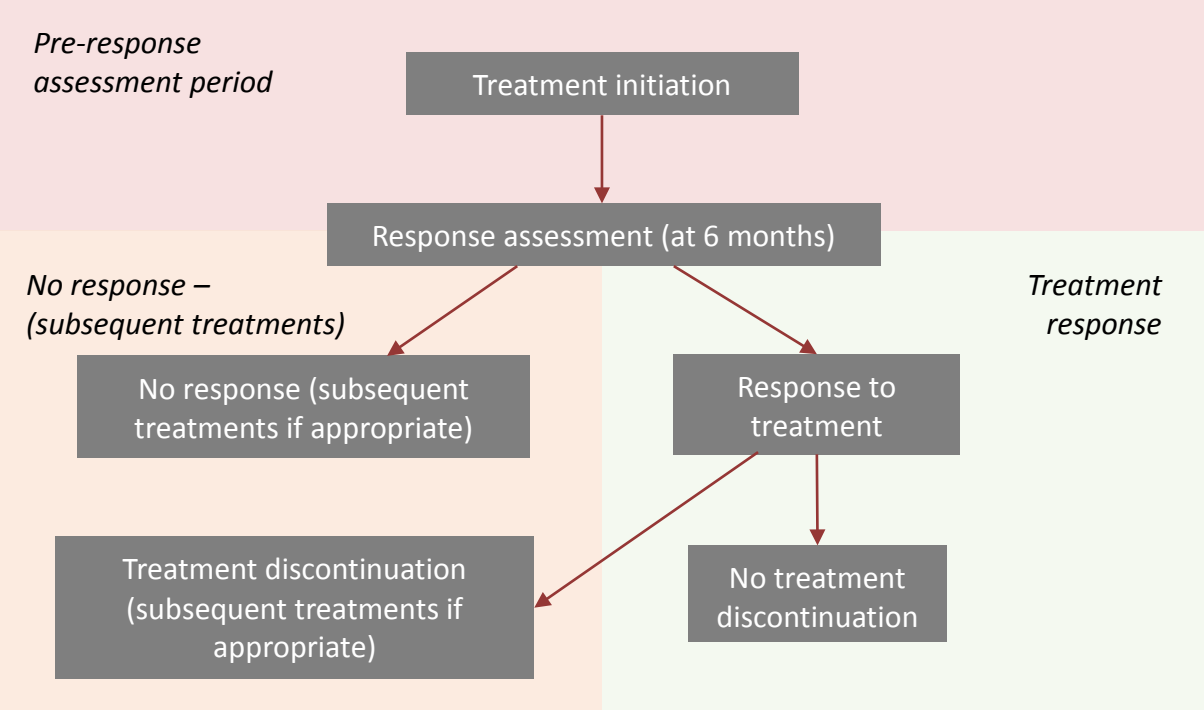


Figure 8: DES – patient flow (copied from Figure 17 of the CS)

In the company's model (summarised in Figure 8), patients are initiated on first-line therapy (i.e. the chosen treatment option) and then all patients are subject to a response check at 6 months. Non-responders are assumed to switch to second-line therapy (including subsequent treatments if appropriate) at this point. Responders remain on their first-line therapy until they discontinue, at which point they switch to second-line therapy. Patients are at risk of dying from all-cause mortality at any point during the model and survival is assumed not to vary by treatment. When BIs are given as the first-line treatment, the model estimates treatment costs and QALYs accrued during BIs by modelling each treatment administration as a separate model event. This allows the frequency of BIs to vary over time. When BIs are given as second-line treatments, the separate treatment administrations are not modelled explicitly and instead annualised costs and QALYs are calculated based on the duration of second-line treatment and the mean number of BI administrations per annum. Therefore, the main events modelled are:

- Response check at 6 months
- Administration of BI (when BI is used as first-line treatment only)
- Treatment discontinuation (in responders only)
- Death from all-cause mortality
- End model due to reaching time horizon (20 years) before death.

The company's model captures the benefit of treatments for IC/BPS through their impact on response rates as patients who respond and stay on treatment are assumed to have higher utility values and lower disease management costs.

The ERG does not understand why the company did not build a state transition model. With the exception of the administration of BIs, the costs and QALYs are determined mainly by the time spent on first and second-line treatment. Therefore, a simple Markov model could have been constructed with health states for:

- Patient on treatment before the re-response check
- Non-responders who have switched to second-line treatment
- Responders remaining on first-line treatment
- Responders who have discontinued and moved to second-line treatment
- Death.

Although the frequency of BIs varies in the first 6 months of the model, it is constant thereafter. Therefore, the varying cost of BIs could have be incorporated simply by having a 6-month cycle length and a different cost in the first cycle. The company claims that the DES structure allows ICSI to be incorporated as a continuous variable. However, ICSI is not implemented in the DES as a continuous variable. Instead average costs and utilities are estimated for responders and non-responders based on their estimated median ICSI scores. This would therefore allow costs and utilities to be easily attributed to the health states listed above. The ERG considers that a state-transition approach would have been more parsimonious but that this does not mean that the DES approach is incorrect.

5.2.3 Population

The modelled population is adults with bladder pain syndrome characterised by either glomerulations or Hunner's lesions with moderate to severe pain, urgency, and frequency of micturition. The ERG considers the overall population to be consistent with the NICE scope[2] and the licensed indication for PPS.[8]

However, the ERG notes that based on the company's presentation of the clinical pathway in Figure 2 of the CS, patients would not receive a cystoscopy and a confirmed diagnosis of IC/BPS until after they had failed to respond to first-line oral therapies including analgesics, antihistamines and other non-pharmaceutical interventions including as dietary and lifestyle advice. Therefore, the population modelled is assumed to be those who did not respond to these initial interventions and the cost-effectiveness results should not be extrapolated to patients earlier in the clinical pathway.

The CS[1] presents cost-effectiveness analyses for two distinct subpopulations;

- Patients able to receive BIs
- Patients for whom BIs are contraindicated or who are unable to tolerate BIs.

The ERG considers it reasonable for the company to present separate estimates of cost-effectiveness for these two subpopulations, as patients unable to receive BIs have different first and subsequent treatment options available.

The ERG is satisfied that it is not necessary for the company to present an analysis of PPS versus BSC in the population able to receive BIs, as offering BIs was considered by the ERG's clinical experts to be established practice in the NHS for this population.

The CS[1] states that all patients are assumed in the model to have a starting age of 45.57 years based on the characteristics of patients in the RCT reported by Sant *et al.* (2003).[17] The company states that 89% of the modelled population are female, but in practice, the sex of the patients is not varied within the cohort, and instead the all-cause mortality rate is calculated as a weighted average of males and females using the proportion who were female (89%) in the Sant *et al.* (2003) RCT.[17] The ERG's clinical experts confirmed that the majority of patients they see with IC/BPS are female and the populations in the included RCTs including Sant (2003) were generally comparable to the UK IC/BPS population. The ERG notes that the error introduced from not modelling males and females separately is likely to be small given that they are assumed to differ only in their life-expectancy and the model is limited to a 20-year time-horizon.

5.2.4 Interventions and comparators

The treatment pathways modelled are summarised in Table 18. In the population able to receive BIs, patients who do not respond to the first-line treatment, or who respond initially but later discontinue, are offered BIs as second-line treatment. It is assumed that these patients cycle through the various commercial BI preparations indefinitely until death or the time horizon (20 years) is reached. It is assumed in the model that sodium hyaluronate is used as the first-line BI. Although the ERG heard from clinical experts that there was no standard order of sequence for trying BIs, the similarity in price of the various commercially available BIs meant that the ERG was not concerned that this assumption had significantly biased the estimates of cost-effectiveness.

Table 18: Modelled treatment pathways

Population	First-line treatment	Second-line treatment given after non-response or discontinuation of first-line treatment
Patients able to receive BIs	PPS given orally as three 100mg doses per day	BIs given every 4 weeks using either sodium chondroitin (2%, Uracyst®; 0.2%, Gepan®) or a combination of sodium hyaluronate and sodium chondroitin (iAluRil®).
	BIs with sodium hyaluronate (Hyacyst®, Cystistat®) given weekly for 4 weeks, then every 4 weeks	BIs given every 4 weeks using either sodium chondroitin (2%, Uracyst®; 0.2%, Gepan®) or a combination of sodium hyaluronate and sodium chondroitin (iAluRil®).
Patients unable to receive BIs	PPS given orally as three 100mg doses per day	BSC
	BSC	BSC

The ERG notes that some clinicians have access to locally prepared BIs which may be lower cost than the commercially prepared BIs. However, these were noted to vary from hospital to hospital and therefore cannot be considered to be part of established practice in the NHS in England. Given that there are licensed commercial bladder instillations available, which the ERG's clinical experts accepted were part of the standard care in the NHS in England, the ERG considered it reasonable to exclude the locally prepared BIs from the economic modelling.

Patients unable to receive BIs who have BSC as their first-line treatment are assumed to continue with BSC regardless of whether they respond as no alternative treatments are available in patients unable to receive BIs.

5.2.5 Perspective, time horizon and discounting

The model takes an NHS perspective. The CS states that PSS costs are not relevant to the decision problem. [1] This was considered reasonable by the ERG.

The base-case analysis uses a time-horizon of 20-years, but also presents a scenario analysis using a lifetime horizon. The ERG considers the use of a lifetime horizon more appropriate for the reference case analysis given that the survival function used to extrapolate time to discontinuation predicts that 18% of patients remain on treatment at 20 years. Therefore, under the company's assumption, the

decision to offer PPS will potentially incur costs that fall outside of their preferred 20-year time horizon, which is inconsistent with the NICE reference case. [38]

The model applies discount rates of 3.5% to both costs and QALYs and is therefore consistent with the NICE reference case.[38]

5.2.6 Sources and assumptions used to inform the model

The key sources used to inform the model are summarised in Table 19. These data sources are discussed in more detail in sections 5.2.7 to 5.2.10.

Table 19*:* Summary of data sources used to inform the company's base-case analyses

Parameter group	**Source**
Patient characteristics (age, gender)	Based on characteristics of trial participants in Sant *et al.* (2003)[17]
Mortality - general population	Derived from interim life tables for England (2015-2017)[42]
BSC response rate	Company's meta-analysis of four RCTs of PPS versus placebo in patients with IC/BPS[14-17].
PPS response rate	RR from company's meta-analysis of four RCTs of PPS versus placebo multiplied by BSC response rate[14-17]
BI response rate	RR from company's meta-analysis of two RCTs in patients with BPS multiplied by BSC response rate[20, 21]
TTD – PPS	Exponential model fitted to observed time to discontinuation data for PPS patients reported in Hanno *et al*.(1997)[27]
TTD – BI	TTD data for PPS used for time to discontinuation of BIs
ICSI scores for responders and non-responders	Mean ICSI scores for responders and non-responders were estimated based on ICSI scores in the PPS arm of RCT by Sant *et al* (2003)[17]
HRQoL	EQ-5D-5L and disease severity measure (ICSI) data collected from a survey conducted by the company of 252 BPS patients [39]. EQ-5D-5L responses valued using crosswalk to EQ-5D-3L UK value set [40]. Utility scores estimated by mapping from ICSI score to EQ-5D [39]. Mapping regression also contained terms for age and recent

Parameter group	Source
	use of BIs with the coefficient for recent use of BIs applied to those receiving BIs in the model.
Disease-related resource use	Company's survey of 252 BPS patients [39]
Disease-related costs	Resource use data from the patient survey[39] were combined with unit costs from the NHS Reference Costs (2017/18)[43] and PSSRU (2017)[44] to estimate disease-related costs as a function of ICSI scores.
PPS drug acquisition costs	Provided by the company
BI drug acquisition costs	Monthly Index of Medical Specialties (MIMS) 2018, NHS Electronic Drug Tariff [45], Company provided Uracyst® cost from manufacturer
Drug administration costs	NHS Reference costs (2017/2018) [43]

The key structural assumptions within the model are as follows:

- The mean placebo response rate from the RCTs of PPS is assumed to apply to patients having BSC
- The effectiveness of first-line BIs with sodium hyaluronate is assumed to be the same as the effectiveness of sodium chondroitin sulphate
- The cumulative rate of response across all subsequent lines of BIs is assumed to be the same as the effectiveness of sodium chondroitin sulphate
- Those responding to either PPS or BI at 6 months are assumed to continue to respond until treatment discontinuation
- Those responding to BSC at 6 months are assumed to continue to responder for 12 months
- Non-responders who have BSC as their second-line treatment are assumed to maintain any changes in utility that occurred during the first 6 months for the rest of the model horizon
- Patients who respond to treatment are assumed to have lower ICSI scores than non-responders and the change in ICSI scores is assumed to be normally distributed.
- The change in ICSI scores for responders and non-responders is the same for all treatments
- Time to treatment discontinuation for BIs is assumed to be the same as for PPS.

These assumptions are discussed in more detail in sections 5.2.7 to 5.2.10.

5.2.7 Treatment effectiveness – response rate

For the comparisons of both PPS versus placebo and BI versus placebo, the trial outcome of GRA has been used to determine whether the patient has received an adequate response to treatment.

The treatment effectiveness of PPS versus BSC is based on a meta-analysis of the four key trials which were conducted in patients with IC/BPS[14-17] (i.e. those with BPS and evidence of either Hunner's lesions or glomerulations). The forest plot for this meta-analysis is provided in Figure 11 of the CS[1] and the fixed effects RR of 2.09 (CI 1.47-2.97) has been applied in the model.

The treatment effectiveness of BI compared with BSC is based on a meta-analysis of two studies[20, 21] in patients with BPS (i.e. no requirement to have evidence of either Hunner's lesions or glomerulations). The forest plot is provided in Figure 13 of the CS[1] and the fixed effects RR of 1.39 (CI 0.89-2.17) has been applied in the model.

An estimate of the GRA baseline response is required in order to generate an estimate of the absolute GRA response rates for each treatment for inclusion in the economic model. The response rate for BSC in the economic model has been estimated by meta-analysing the response rates in the placebo arms of the 4 RCTs which compared PPS to placebo in the IC/BPS population[14-17] (16%, 95%CI 0.12-0.21, see Figure 12 of the CS).

The ERG notes that the estimates of relative treatment effect for response to treatment for PPS versus BI, that inform the estimate of incremental cost-effectiveness for PPS versus BI, are dependent on the company's simple unadjusted indirect comparison between PPS and BI. In practice, this means that the rate of response in the PPS group is equal to the rate of response for BSC (16%) multiplied by the RR for PPS vs placebo from the meta-analysis (2.09) to give a response rate for PPS of 33%. Similarly, the response rate in the BI group is the response rate for BSC multiplied by the RR for BI vs placebo (1.39) to give a response rate for BI of 22%. Therefore, the effective RR in the model for PPS versus BI based on the indirect comparison is 1.50 (=2.09/1.39). A critique of the company's systematic review and meta-analysis, which inform these estimates of relative treatment effect, including the ERG's concerns regarding the indirect comparison, is provided in Section 4.

The ERG has a concern with the relevance of the estimate of GRA baseline response (16%) given that the company has stated that the RCTs of PPS and chondroitin sulphate reported high response rates in their placebo arms (see page 104 of the CS). The estimate of the GRA response rate expected in an untreated population in clinical practice can come from sources other than clinical trials, for example, registries and expert opinion. Clinical advisors to the ERG suggested that 20-30% of patients with IC/BPS (especially those with milder symptoms) would be expected to report an improvement in

clinical practice in the absence of treatment with either bladder instillations or oral PPS. On the other hand, the CS states that the high placebo response rates observed in the clinical trials, estimated by the company to be 16% (CS, Figure 12), are unlikely to be observed in clinical practice.[1] As part of the clarification process, the ERG asked the company to state the expected GRA response for patients receiving standard of care in clinical practice (see clarification question A18). The company replied:

> *"The main change in clinical practice since the trials were conducted is that standardised, commercially-available bladder instillations are now routinely used in the treatment of BPS. As noted in our submission, it is difficult to disentangle the effect of placebo in the clinical trials of PPS and BIs. We are unaware of any contemporary data reporting the 'response' to standard of care i.e. initial treatments (e.g. pain management, etc). In our analysis, we have adopted a highly conservative approach of assuming the placebo effect would be observed in clinical practice for a year for patients not receiving PPS or BIs, even though this response is likely to be due to participation in the trials. Please note that this assumption is likely to underestimate the effectiveness of PPS,"* (clarification response, question A18).[1]

The ERG notes that assuming a GRA response rate for BSC that is similar to the placebo response in the clinical trials (that is believed to be higher than expected for an untreated population as a consequence of participating in a clinical trial) is likely to benefit the company rather than being conservative. This is because the absolute effect of PPS is estimated by applying a relative risk and the absolute difference becomes greater with increasing baseline response. The exact impact on the ICER will depend on whether incremental costs and QALYs vary at the same rate when the baseline risk is varied. The ERG considers that the true response rate for patients receiving BSC in clinical practice is uncertain. The ERG therefore conducted exploratory analyses to examine the impact on the ICER of raising and lowering the response rate in the BSC arm of the model (see Section 5.3).

In addition to the ERG's concerns regarding the lack of an appropriate estimate of response rate for patients receiving BSC, the ERG also notes that the company makes different assumptions in the model about the durability of the response achieved for patients in different arms of the model. The company argues that the response rates observed in the placebo arms of the RCTs would be unlikely to be observed in patients receiving BSC in clinical practice because they are "*likely to be a result of participating in the clinical trial.*" To account for this in the model, the company limited the benefits for responders in the BSC arm to the first 12 months of the model. In contrast, the responses achieved in patients receiving PPS or BIs as first-line treatment are assumed to persist until treatment is discontinued.

The ERG considered that it was inconsistent to assume that all of the responses observed in the PPS and BI arms of the RCTs were durable, in that they would persist until treatment ceased, but all of the responses observed in the placebo arms of the RCTs were not durable and would cease at 12 months.

If the response rate observed in the placebo arms of the RCTs was related to the experience of being enrolled in a clinical trial, then it may also apply to a proportion of the patients who responded in trial arms receiving either PPS or BI. If the response rate in the placebo arm is related to the fact that patients may enrol in the trial when experiencing a flare-up in their symptoms, which resolves naturally over the course of the trial (i.e. regression to the mean), then again, it does not seem reasonable to assume that this response is time limited in patients receiving BSC, but continues indefinitely in those receiving PPS or BI. RCTs are designed to provide an unbiased estimate of the relative treatment effect. It is this relative treatment effect that should inform the differences in outcomes between treatments within the economic model. However, the company's assumption that benefits are limited to 12 months in patients responding to BSC introduces a difference in the model that is separate from the relative treatment effect measured in the trial. The ERG does not consider that this is reasonable given that the company has provided no evidence to demonstrate that the durability of response differs in patients receiving BSC compared to those receiving either PPS or BI.

5.2.8 Treatment effectiveness – extrapolation

In the PPS and BI arms of the model, patients who have responded after 6 months of treatment are assumed to continue receiving the full treatment effect until they discontinue. The time-to-discontinuation survival function is based on data from Hanno *et al.* (1997)[27] which has a maximum follow-up of 10 years. An exponential survival function is then used to extrapolate discontinuation rates over the remainder of the model. The median time to discontinuation in the company's model, based on their preferred parametric survival function, is 7 years with 18% of patients estimated to still be on treatment after 20 years. The effectiveness of PPS and BIs has therefore been extrapolated for some patients for up to 20 years in the company's base-case analysis. This is in contrast to the RCTs having a maximum of 6-month follow-up for assessment of response based on GRA. The ERG is concerned that there is a lack of data on the long-term efficacy of PPS despite the drug having being available in Canada, Australia and the US for over 20 years. Whilst some data on efficacy up to 36 months are provided in the CS, these are from an observational study which is poorly reported and as such are difficult to interpret. (see Section 4.2.6)

In addition, the data on discontinuation are based on a study in patients treated with PPS, but the same survival function for time to discontinuation is also applied in the model to patients receiving BIs as first-line treatment. No evidence is provided to support the assumption that rates of discontinuation would be the same for BIs and PPS. Given that these treatments vary substantially in their mode of

administration, it is possible that the rate of discontinuation may differ substantially. For example, given the invasive nature of BIs, it is possible that patients may stop and restart treatment according to the severity of their symptoms, resulting in fewer BIs per annum than predicted by the model. One of the ERG's clinical experts also stated that they considered it unlikely that 18% of patients would still be taking BIs at 20 years given that BIs are an invasive treatment. Therefore, the ERG considered that the discontinuation rates for BIs lacked face validity.

Patients starting second-line treatment with BIs are assumed to continue on second-line BIs for the rest of the model horizon. Therefore, the effectiveness of BIs compared with BSC in second-line patients has been extrapolated for up to 20 years in the base-case analysis. Patients are assumed to cycle through the various BIs available until they achieve a response. The response, or lack of response, to each subsequent BI is not modelled explicitly. Instead, the costs and utilities in patients having BIs as second-line treatment are based on the mean response rate to BIs when used as first-line treatments. No evidence is provided to support the assumption that the cumulative response rate achieved over numerous lines of subsequent BIs will be the same as the response achieved during first-line BI treatment. It is also unclear whether patients are likely to cycle through second-line BI treatments indefinitely (including treatments that they have previously failed on, as claimed on page 34 of the CS), or whether some would transition to BSC over time. This is important given that the costs and utilities applied to those on BSC differ to those remaining on second-line BIs. However, it is difficult for the ERG to estimate the size and direction of any potential bias, given that response to second-line BIs and discontinuation from second-line BIs is not explicitly captured in the company's model structure.

5.2.9 Health-related quality of life

The model estimates ICSI scores for responders and non-responders using data from the PPS arm of the RCT by Sant *et al.* (2003[17]). Data from a patient survey of 252 BPS patients were used to map from ICSI scores to utilities as measured by the EQ-5D.[39] The EQ-5D-5L responses from the patient survey were mapped to EQ-5D-3L responses using the algorithm reported by van Hout *et al.*[40] and the UK valuation set for EQ-5D-3L were applied. A mapping regression was then fitted to estimate EQ-5D-3L utilities as a function of ICSI, age and a term that captured prior use of BIs (see Table 43 of the CS). The different ICSI scores predicted for non-responders and responders from the PPS arm of the RCT by Sant *et al.* is therefore used to determine the utility gains associated with a response to treatment in the model. The regression coefficient for the term "received a bladder instillation in the previous 6 months" was applied to all patients having BIs in the model. This included those having first-line BIs before the response check, those responding to first-line BIs and those having BIs as a second-line treatment after either first-line PPS or first-line BIs.

The utility scores applied in the model are summarised in Figure 9 according to whether patients respond or do not respond to first-line treatment and according to whether they discontinue following response (see Table 48 of the CS[1] for the numerical values). We have illustrated the scenarios in Figure 9 assuming that patients who discontinue do so at exactly 7 years as this is the median time of treatment discontinuation; however, in the model patients can discontinue at any time from 6 months to 20 years according to the survival function for time to discontinuation.

Figure 9:

It can be seen from Figure 9 (panels A and B) that patients who respond to PPS or BIs are assumed to continue to benefit from improved HRQoL until they discontinue treatment, but may continue to benefit for the full model horizon if their time to discontinuation is sampled to be greater than 20 years; this occurs in 18% of responders. In contrast, patients who respond on BSC, are assumed to have a HRQoL benefit that lasts only from the response check at 6 months to 1 year after the start of the model (NB: alternative scenarios are provided in the CS[1] where the treatment effect stops immediately after the 6-month response check or at 5 years).

It can also be seen from Figure 9, that in the scenario where patients cannot have BIs (panels C and D), patients who do not respond to either PPS or BSC experience some HRQoL improvement due to an assumed improvement in ICSI scores in non-responders compared with baseline. The ERG notes that the utility score of non-responders having BSC in this scenario ([redacted]), is [redacted] than the average utility score achieved by patients having BIs as second-line treatment ([redacted]) even though 22% of these patients respond to second-line BIs. This inconsistency is being driven by the utility decrement associated with receiving BIs which results from the regression coefficient for having "received a bladder instillation in the previous 6 months". The ERG notes that only 53% of the undiscounted QALY gain for PPS versus BIs is accrued in patients who responded on PPS but would not have responded on BIs. The remainder is due to differences in QALYs that result from time spent on first-line BIs due to the application of the regression coefficient for having "received a bladder instillation in the previous 6 months".

The ERG does not understand the clinical rationale for there being a utility decrement associated with having previously received BIs. In response to a request for clarification, the company stated "*Bladder instillations are an invasive and uncomfortable procedure, and have been associated with adverse effects. Clinical experts confirmed the likelihood of reduced quality of life with bladder instillations, highlighting in particular the potential for an increase in urinary tract infections*".[1] However, the ERG is concerned that the difference in utility detected in the patient survey[39] may reflect differences in patient characteristics in the survey population between those who have recently used BIs and those who have not recently used BIs. In this case it would not be appropriate to apply it only to those having BIs in the model as it is related to the population and not the current treatment. Furthermore, although the survey did ask about oral medications, the number reporting use of oral PPS was considered by the company to be insufficient to robustly include a covariate for PPS treatment in the mapping model.[39] Therefore, it is not possible to know if there is a similar decrement associated with taking PPS that could not be detected in the survey.

[redacted]

[redacted] The ERG decided to use the regression including age and ICSI score but not BI usage in their base-case analysis as the ERG was not satisfied that [redacted]

The EQ-5D values for responders and non-responders are based on estimates of median ICSI scores for responders and non-responders. These have been calculated by assuming that ICSI scores in the PPS arm of the RCT by Stunt *et al.* (2003)[17] are normally distributed and that all patients who respond have ICSI scores that are lower (i.e. better) than all patients who do not respond (see Figure 18 of the CS). The ERG notes that the company were unable to provide any data to support these because they do not hold any relevant patient-level trial data (see company response to clarification question B5).[1] Based on these assumptions, the ICSI score for the median responder and the median non-responder was calculated from the normal distribution of the ICSI scores.

The ERG has concern with the assumptions made when relating GRA response to ICSI. The company has effectively assumed a step function such that all patients who have a change from baseline to Week 24 of greater than (approximately) -4.1 in ICSI are considered as non-responders and all patients who have a change from baseline to Week 24 of less than (approximately) -4.1 in ICSI are considered as responders and that this applies irrespective of treatment (CS, pages 104-105). The ERG suggests that it is unlikely that such a dichotomy according to baseline ICSI will be true or that there will be no treatment effect. In addition, the ERG has additional concerns with the analysis as implemented by the company:

- The company assumes that the underlying model for the ICSI data is a normal distribution without providing any justification for this.

- The assumptions regarding the 33.9% (20/55) GRA response rate, and the sample mean and standard deviation for the change from baseline to week 24 ICSI from Sant *et al.*, 2003[17] ignore uncertainty in their estimates.
- The absolute central estimates of ICSI response for non-responders and responders were estimated by adding the median estimates (-1.11 for non-responders and -5.85 for responders) to a baseline response. The ERG suggests that means would be more appropriate than medians, which it estimates to be -6.33 and -0.76 approximately for responders and non-responders, respectively. The ERG notes that when the company did this it had a minimal impact on the ICER (see response to clarification question B6),[1] although this scenario analysis did not address the ERG's concern that all responders are assumed to have higher ICSI scores than all non-responders.

The same ISCI scores were assumed to apply to responders in the BI and BSC arms of the model. This was done to ensure that the benefits received by responders compared to non-responders were consistent across the model. The ERG considers that whilst this is a pragmatic approach which simplifies the model inputs, it is implausible to assume that all responders have the same degree of response. Given that the company states in its rationale for using a DES structure, "*As well as considering response, the DES allowed the incorporation of evidence on likely magnitude of response based on a continuous scale"*, it seems fairly crude to then reduce the model to one based on a binary response / no response outcome, with identical benefits assumed for all responders.

The ERG noted that utility values in non-responders were generally [redacted] than the utility values pre-response assessment which were based on baseline ICSI scores. It may be reasonable that there is some [redacted] in ICSI scores in those patients who do not have a sufficient reduction in symptoms to be classed as a responder, and therefore there is a predicted [redacted] in utility for non-responders at 6 months. However, the utility values for non-responders after 6 months of treatment are being applied in the model to patients who discontinue treatment after being classed as non-responders. Therefore, it would seem reasonable that [redacted] and it would be more appropriate to assume that non-responders return to their baseline utility value unless they switch onto another active treatment.

The ERG noted that the pre-response assessment utility value based on baseline ICSI scores is being applied as a constant value during the first 6 months. This may be considered conservative if there is some symptomatic benefit from the day treatment is started. However, the ERG's clinical experts advised that the treatment effect is known to build slowly over time for both PPS and BIs, and

therefore the company's assumption that utility values for responders are not updated until 6 months is reasonable.

Finally, the ERG notes that the utility values in the model are not adjusted for age-related utility decrements. The ERG notes that the highest utility value applied in the model [redacted]. However, the ERG considers that it would have been preferable to either have capped the utility values at the values for age-matched general population norms in the company's lifetime horizon or to have estimated a proportional utility decrement relative to general population norms which could then be applied to age-related general population norms in the model. Furthermore, the application of constant utilities across time appears to contradict the evidence from general population studies that utilities generally decline with age. The ERG accepts that it is technically more difficult to apply age-related utility decrements within a DES model than within a state transition model, because utility values can only be updated at the point that events occur in a DES rather than every cycle in a state transition model. However, the ERG notes that age-adjustment of utilities can be achieved using either dummy events which update the utility values at regular intervals (say every 5 years), or by assuming a linear change in utilities between the previous and the current event and using this assumption to estimate the average utility in the period since the last event.

5.2.10 Resources and costs

The costs included in the model are summarised in Table 20 and the costs over time for patients having different model trajectories are plotted in Figure 10. It should be noted that the graphs in Figure 10 illustrate the equivalent costs per annum (i.e. the actual costs accrued in the first 6 months are doubled to see their size relative to costs accrued per annum in later periods) and the example for a patient who discontinues assumes that they do so at exactly 7 years whereas patients can discontinue anytime from 6 months to 20 years and the exact time varies from patient to patient.

Table 20: Summary of costs applied in the company's model

Description of cost	Annual costs*	Source
Disease-related costs for pre-response assessment (PPS/BIs/BSC)	[redacted]	Regression for costs as a function of ICSI based on resource use reported in patient survey[39] combined with NHS reference costs[45]
Disease-related costs in responders (PPS/ BI / BSC)	[redacted]	
Disease-related costs in non-responders who have switched to second-line BIs (a proportion of whom are assumed to respond to second-line BIs).	[redacted]	
Disease-related costs in patients on BSC after non-response to either PPS or BSC	[redacted]	
PPS drug treatments (pre-response and responders up to discontinuation)	[redacted]	Company[1]
BI as first-line treatment pre-response assessment - 9 administrations in 6 months with acquisition cost of £88.03 for BI cost of £183.37 for administration	[redacted]	MIMS[47] for list prices of medical devices and NHS reference costs[45] for administration
BI as first-line treatment in responders - 13 administrations in a year with acquisition cost of £88.03 for BI cost of £183.37 for administration	£3,535	
Drug costs for BIs as first-line p.a. 13 administrations** in a year with acquisition cost of £86.14 for BI cost of £183.37 for administration	£3,510**	

* except first-line BI pre-response check which is given as per 6 months

** updated by company in revised model post clarification process

Figure 10:

The three main types of resource use incorporated in the model are: (i) acquisition costs for PPS and BIs; (ii) administration costs for BIs, and (iii) disease-related costs. The latter is assumed to be related to disease severity as measured by the patient's ICSI scores. Data from the patient survey were used to estimate the relationship between costs in the previous 6 months and ICSI scores.[39] This relationship was used to estimate annual costs for responders and non-responders using the ICSI scores previously calculated for estimating utility based on ICSI. Again, as when calculating utilities, age was also included in the regression for costs, but the costs were calculated based on patient age at the start of the model and were not updated as patients aged in the model. The main difference between the approach used for utility and that used for resource use was that no explanatory term related to previous BI use was included in the regression linking ICSI scores to health care costs.

It should also be noted that the company attempted to remove any double counting of costs directly related to interventions. However, in the survey, patients were asked separately about hospital visits and treatments received without any information being gathered on whether the resource use was related to treatments received.[39] Therefore, it is possible that treatment-related resource use has not been adequately excluded as intended. This may mean that disease-related costs are over-estimated in the model.

In calculating the overall cost in the previous 6 months from the survey results, the company applied HRG costs to the resource use data.[39, 45] In several cases, it was unclear how the various HRG costs were selected and why other values were not applied. For example, the HRG cost applied for hospital admissions is the weighted mean across elective, non-elective and day-case admissions for that HRG code. In their response to clarification question B25, the company stated that, [redacted] The data collected [redacted] do not appear to have been used to inform the average HRG cost for hospital admission.

The ERG asked their clinical experts whether patients with poorer disease control, and therefore higher ICSI scores would be likely to incur greater resource use and whether the types of resource use reported in the patient survey (Table 49 of the CS)[1] were typical based on their experience. The ERG's clinical experts agreed that patients with poor symptom control may be more likely to access NHS services, but these were likely to be outpatient services rather than inpatient admissions or emergency department (ED) attendances. One clinician noted that the incidence of GP appointments

may increase if the patient does not have easy access to outpatient services. The ERG noted that whilst [redacted] of the costs in Table 49 of the CS[1] were related to outpatient visits, the proportion relating [redacted] respectively. The ERG was not convinced that these costs were necessarily related to IC/BPS. In particular, the inclusions of costs [redacted]. The ERG was concerned that no attempt had been made to estimate the costs in patients with IC/BPS relative to matched controls without IC/BPS. The ERG's concern is that the disease-related costs have been overestimated as not all of the resource use reported in the survey is attributable to IC/BPS.

A comparison of the HRG costs applied by the company (in Table 49 of the CS) and those preferred by the ERG is provided in Appendix 1. The ERG's concern here is that disease-related costs may have been overestimated in the model, but based on the comparison presented in Appendix 1 any overestimation is likely to be small at around 6% of the total cost estimated by the company.

None of the individual HRG costs are applied directly in the model. Instead the model inputs are based on the outputs of the regression, with the regression coefficient for ICSI score being key in determining the difference in costs between treatment arms. The ERG was not able to revise the HRG costs and update the regression analysis to re-estimate the regression coefficient for ICSI score without access to the full patient survey data. Therefore, it was not possible for the ERG to quantify the extent of any bias introduced from the choice of HRG costs. Instead the ERG explored whether the relationship between ICSI scores and resource use was an important determinant of cost-effectiveness by removing the dependence of resource use on ICSI scores in a scenario analysis (see exploratory analysis 3 in section 5.3).

The drug costs for PPS were provided by the company. The acquisition cost for BIs was based on the mean cost for two preparations of sodium hyaluronate, weighted using their market share. Prices for bladder instillations were generally taken from MIMS,[47] but these were cross-checked by the ERG with the NHS Electronic Drug Tariff (Feb 2019). The exception was Uracyst where the CS[1] stated that the price was sourced from the manufacturer and consequently this price could not be verified by the ERG.

No costs associated with monitoring or administration were applied to patients receiving PPS. The ERG's clinical experts generally agreed that patients on oral PPS would not require more intensive monitoring than patients not having oral PPS, as patients with IC/BPS would generally be seen every 3 to 6 months in clinic irrespective of whether they were on oral PPS treatment.

The cost for administration of a BI was based on the HRG for "Introduction of a therapeutic substance into the bladder". The company applied the mean cost across all types of care, rather than applying the

specific cost for day case or outpatient procedures. The clinical advisors to the ERG stated that BIs are commonly given as outpatient procedures, although one also stated that they were sometimes done as day-case procedures. The costs for the relevant HRG code are £151 for outpatient and £223 for day-case procedures, with a weighted mean of £185, which is close to the cost applied in the company's model. However, given that the majority of the clinical experts reporting BIs being administered as outpatient procedures, the ERG conducted a sensitivity analysis applying this cost (see ERG sensitivity analysis 4 in section 5.3).

In their base-case analysis, the company assumed that no patient would self-administer their BIs, although a self-administration rate of 10% was explored in a scenario analysis. The ERG asked their clinical experts what their experience was regarding patients self-administering BIs. There appeared to be significant variation in usual practice, with two clinical experts suggesting it was not routine practice for patients to self-administer and two reporting that a high proportion (64% and 80%) are able to self-administer once they have been trained to do so. The ERG therefore conducted a scenario analysis exploring the impact of high rates of self-administration of BIs on the cost-effectiveness of PPS relative to BI (see ERG sensitivity analysis 5 in section 5.3).

The company assumed that first-line BIs would be given every week for 4 weeks followed by every 4 weeks thereafter. They also conducted a sensitivity analysis exploring the impact of assuming administration every 6 weeks after the first month. The ERG asked their clinical experts what the frequency of BIs was in routine clinical practice. There did not seem to be a consistent protocol for the frequency of BI administration, although all of the clinical experts agreed that the frequency of administrations would be weekly initially and would then reduce. Some stated that the interval between instillations would be dependent on the maximum interval the patient could tolerate and in some patients the interval could be as long as 2-3 months depending on response. It was also noted by several clinical experts that patients may discontinue once their symptoms are under control and then they may return several years later after experiencing a flare-up of symptoms. Overall, the ERG considered that treatment frequency for BIs was likely to be higher than 6-weekly in the first year of treatment, but the average frequency was likely to be lower than 4-weekly in the long-term. The ERG decided to implement 6-weekly administrations of BIs from 1 years onwards for first-line BIs and for all patients receiving second-line BIs in their base-case scenario (see ERG exploratory analysis 2 in section 5.3).

Although the NICE scope for this STA explicitly states that the economic modelling should include the costs associated with diagnostic testing for glomerulations or Hunner's lesions in people with bladder pain syndrome who would not otherwise have been tested, the company's model does not incorporate any costs for cystoscopy because the CS argues that cystoscopy is carried out in all

patients as part of the standard diagnostic pathway.[1] The ERG is satisfied that this is reasonable based on the advice provided by their clinical experts who stated that IC/BPS is generally a 'diagnosis by exclusion', and cystoscopy is routinely used to exclude other conditions with similar symptoms before the diagnosis of IC/BPS is made.

The ERG notes that no adverse events are included in the economic model although it is stated in several places in the CS that BIs are associated with UTIs (see pages 34-36 of the CS). The ERG asked their clinical experts whether UTIs were likely to significantly impact either costs or HRQoL and were reassured that UTIs associated with BIs were usually easily avoided or easily treated if they occurred. The ERG considered that the omission of AEs from the model was unlikely to have significantly biased the estimates of cost-effectiveness.

5.2.10 Time to treatment discontinuation

The ERG noted that in the model, some patients are still on first-line treatment at 20 years and patients are able to stay on second-line treatment indefinitely. The ERG asked their clinical experts about the likely rate of treatment discontinuation from treatment. In general, there was agreement that some patients would come off treatment after a period of successful response but others would need long-term treatment for IC/BPS. One clinical expert stated that patients generally do not stay on treatment for 10-15 years. One clinical expert noted that it would be unlikely for 18% of patients to remain on the same treatment for 20 years, as predicted by the company's base-case model. Based on these responses, the ERG is concerned that the model may overestimate lifetime treatment costs for patients.

The ERG reviewed the study by Hanno *et al.*(1997)[27] that was used by the company to determine time to treatment discontinuation for both PPS and first-line BIs. Patients in this open-label "physician's usage" study had to provide data and receive medical assessments every 3 months. They also had to pay for the medication themselves. Although Hanno *et al.* (1997) state that the minimum duration of treatment was 3 months and the maximum was 35 months, this appears to relate only to patients included in the efficacy assessment.[27] Data on treatment discontinuation in Table II of Hanno *et al*, appear to be provided for all subjects with follow-up from 0 to 60+ months, with the study described as having run from 1986 to 1996. [27]

In response to a request for clarification, the company provided additional information on the dataset extracted from Table II of Hanno *et al.*(1997)[27] and used in the company's survival analysis (see company response to clarification questions B10 to B12).[1] The number of patients known to have discontinued in the company's dataset matched the sum total of those reporting their reason for

discontinuation as being "adverse event" or "lack of efficacy" (column E of Table 21). This was less than the total number known to have discontinued (column D of Table 21).

The ERG did not agree with the company's interpretation of the data presented by Hanno *et al.* (1997)[27] The ERG considers that it would have been more reasonable to include all patients known to have discontinued (column D of Table 21) when estimating the survival function for time to discontinuation. Furthermore, the ERG noted that the totals given for all reasons in Table II of Hanno *et al.*, including the "other" category, did not add up to the total number of discontinuations, suggesting that data on the reasons for discontinuation were incomplete. The ERG therefore did not believe that it was reasonable to allocate some patients recorded as having discontinued to be censored based on their reason for discontinuation. In addition, the company assumed a discontinuation time of 90 months for those reported to have follow-up of 60+ months. The ERG preferred to assume that these patients were censored at 60 months as their exact time of discontinuation is not known.

The company's analysis excluded patients who discontinued in the first 6 months of treatment as they intended to estimate time to discontinuation from the response check at 6 months. The ERG was satisfied that it was reasonable to exclude these patients as the time to treatment discontinuation survival function is applied only from 6 months in the model (these patients have been excluded from Table 21 accordingly). However, the ERG noted that the dataset used in the company's survival analysis used time reported from starting treatment rather than time from completing 6 months of treatment.

The study separates those patients who have not formally discontinued into active and inactive patients, with inactive patients being those that did not have any shipments of the drug in the last year of the study. The company's analysis assumes that both active and inactive patients are censored at the end of their study participation. This seems somewhat inaccurate as inactive patients have discontinued in the sense that they have stopped receiving shipment of the drug. However, due to the poor reporting in Hanno *et al.*(1997)[27] it is difficult to determine how to categorise inactive patients in the survival analysis. The ERG therefore believes that the discontinuation data from Hanno *et al.* (1997)[27] should be interpreted with caution.

The ERG generated an alternative survival data set from the data presented by Hanno *et al.*,[27] with the following changes: time measured from 6 months; all patients recorded as discontinuers included as "failures" in the survival analysis (column D of Table 21) and all other patients categorised as being censored at their longest follow-up (column A minus column D of Table 21), including those whose discontinued after 60 months (column A of the last row of Table 21). Although the ERG prefers this

interpretation of the data from Hanno *et al.*(1997) [27] the study is reported poorly and the correct interpretation is unclear.

Table 21: Summary of discontinuation data from Hanno *et al. (1997)*[27] restricted to those with at least months of study participation

Length of participation in months*	Total patients	Active	Inactive	All discontinued	Discontinued due to adverse event or lack of efficacy	Other reason for discontinuation**
Column indicator	A	B	C	D	E	F
6-12	353	83	15	255	129	108
12-18	166	46	10	110	43	57
24-36	116	37	3	76	33	37
36-48	149	63	8	78	30	41
48-60	88	40	6	42	12	22
60+	67	38	3	26	9	15

* interpreted by the ERG to mean maximum follow-up for that individual (ERG has excluded the data from Hanno *et al.*(1997)[27] for patients who participated for less than 6 months)

** the ERG has combined those reporting death, failed to return or "other" as their reason for discontinuation in column F

Table 22 shows the regression parameters for the company's survival analysis with the corresponding Kaplan-Meier data and fitted survival functions shown in Figure 11. The ERG re-analysed the data using their preferred dataset using STATA(version 15.0)[48] using the STATA function 'stset, dist()' for commonly used probability distributions (see Appendix 4 for details). The regression coefficients for the ERG's survival analysis are provided in Table 23 and Figure 12 shows the Kaplan-Meier data and fitted survival functions. The scale parameter for the exponential distribution (the company's preferred distribution) was 0.0074 when using the company's dataset (Table 22) and 0.0229 when using the ERG's preferred dataset (Table 23). Therefore, the rate of discontinuation was approximately 3 times higher when using the ERG's preferred dataset and the company's preferred model. The ERG explored the impact on the ICER of this higher discontinuation rate (see ERG's exploratory analyses section 5.3).

Table 22: Regression parameters for company's survival analysis of time to discontinuation data

Survival function	Scale	Shape	AIC	BIC	Mean time (months)	Median time (months)
Exponential	0.0074	NA	1556.23	1561.20	135.14	93.67
Weibull	0.0047	/ln_p= 0.1132	1553.26	1563.21	114.95	86.37
Gompertz	0.0093	/gamma= -0.0095957	1547.58	1557.52	NE	130.57
Loglogistic	4.4193	/lngamma= -.2789774	1523.13	1533.08	285.05	83.04
Lognormal	4.4347	/ln_sig= 0.2376	1479.71	1489.66	188.44	84.32

NA, not applicable; NE, not estimable

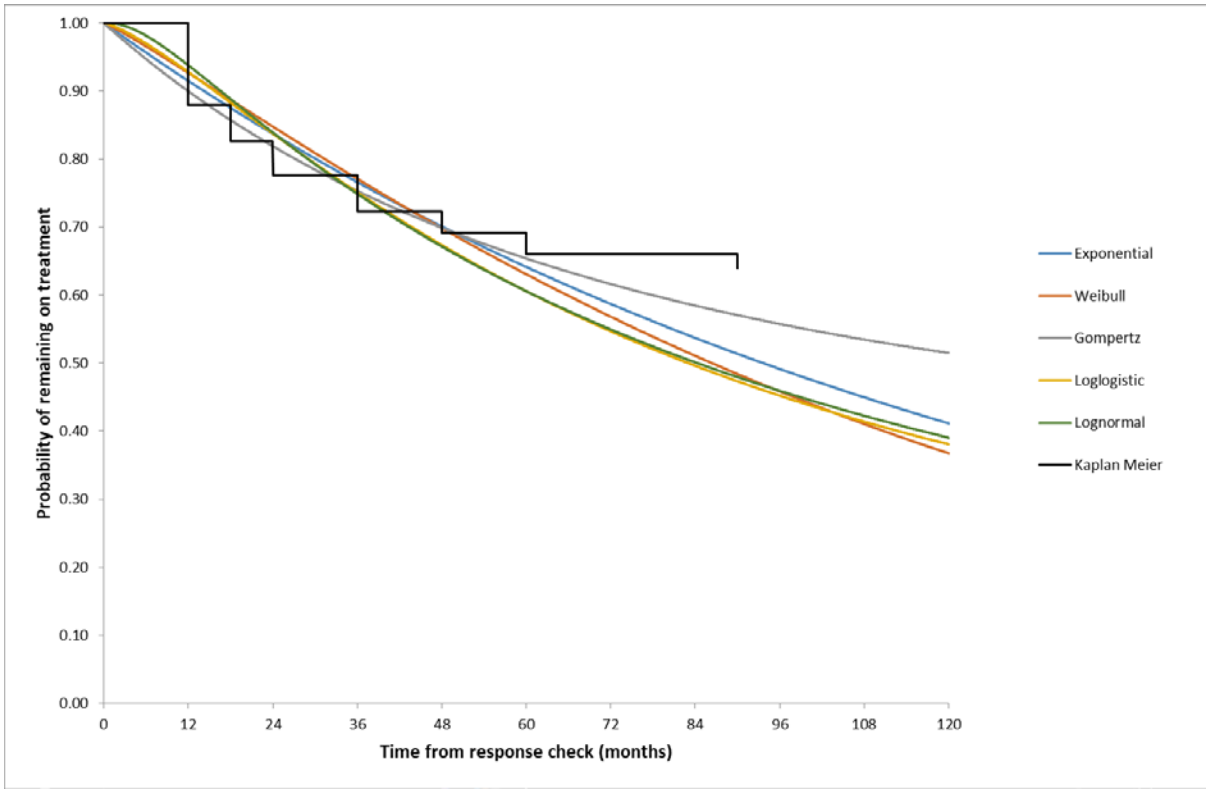


Figure 11: Company's time to treatment discontinuation parametric functions and Kaplan-Meier, reproduced by the ERG

Table 23: Regression parameters for ERG's survival analysis for time to discontinuation

Survival function	Scale	Shape	AIC	BIC	Mean time (months)	Median time (months)
Exponential	0.022874	NA	2659.28	2664.25	43.72	30.30
Weibull	0.021829	ln_p= 0.01329	2661.12	2671.07	43.32	30.34
Gompertz	0.02966	/gamma = -0.0143	2634.45	2644.39	NE	28.44
Loglogistic	3.273407	/ln_gam = -0.29674	2578.45	2588.39	85.38	26.40
Lognormal	3.302363	/ln_sig = 0.18185	2521.39	2531.34	55.80	27.18

NA, not applicable; NE, not estimable

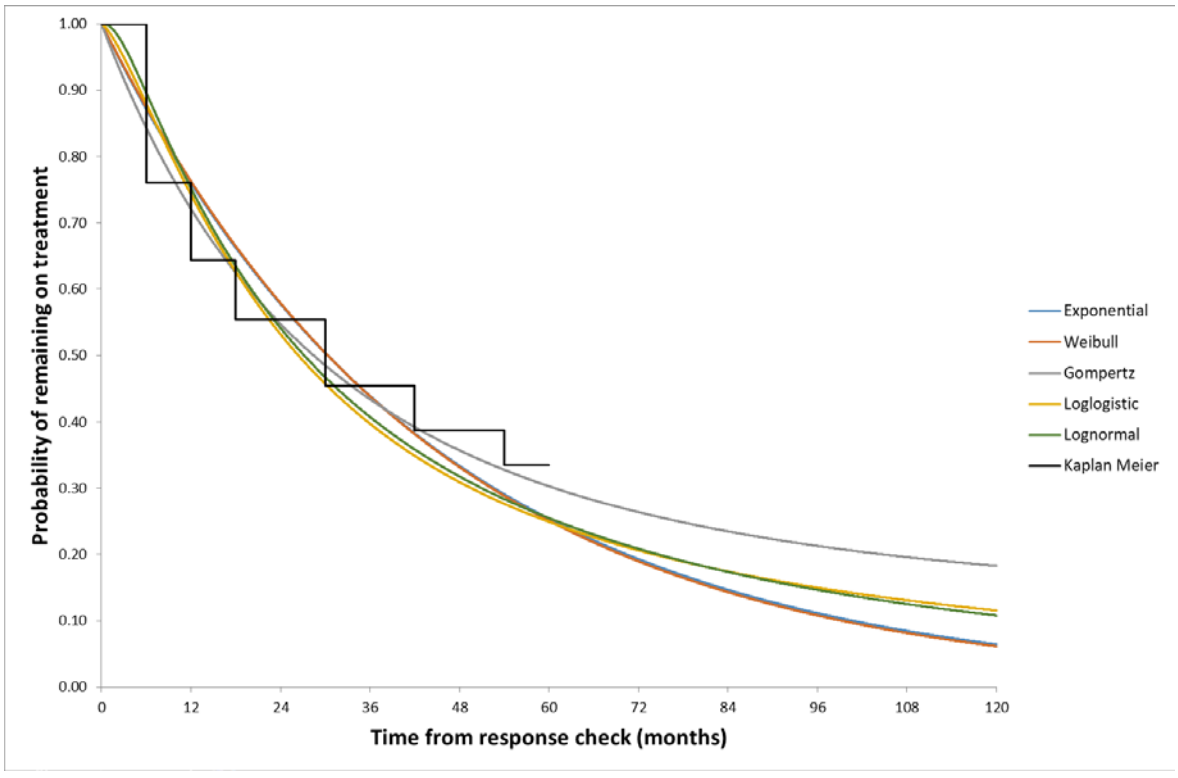


Figure 12: Time to treatment discontinuation parametric functions based on the ERG's preferred interpretation of the data from Hanno *et al.* (1997)[27]

The company's base-case uses the exponential survival function for the time to discontinuation (Figure 11). The ERG notes that the log normal distribution has lower AIC and BIC values in both the ERG's (see Table 23) and the company's analysis (see response to clarification question B13 and Table 22).[1] The log normal distribution predicts a longer mean time on treatment and a shorter median time on treatment than the exponential distribution, reflecting higher discontinuation rates initially which reduce over time. The ERG believed this better reflected the view of the clinical experts: that discontinuation rates would be high initially as some patients achieved resolution of their symptoms and came off treatment, but that discontinuation rates would fall over time, with a subset of patients staying on treatment long-term. The ERG explored the impact on the ICER of using the log normal distribution for time to treatment discontinuation in their exploratory analyses (see Section 5.3).

5.2.11 Mortality

No survival benefit is assumed in the model and mortality risks are constant between arms. The company calculates a normally distributed life expectancy from the data provided in the ONS life-tables.[42] The ERG were not satisfied with their explanation regarding how the SD for this distribution was calculated from the data provided in the ONS life-tables despite the company providing further details in response to a request for clarification (see responses to clarification questions B32, B33 and B34).[1]

The ERG was not satisfied with the company's assumption that life expectancy at a given age would be normally distributed. The company argued in their response to the clarification request that their rationale for using a normal distribution for life-expectancy was based on the example model provided with the NICE DSU's Technical Support Document on patient-level simulation (TSD15).[1, 49] Whilst such an assumption is used in the simple example model provided with TSD15, it is not recommended as the best method for sampling life expectancy from life tables within TSD15.[49] The ERG's preferred method is to use the OLS life tables data to generate an empirical distribution for life expectancy dependent on the starting age. However, because survival in the model is assumed to be identical for patients receiving PPS, BI and BSC, the ERG did expect any significant bias to have been introduced by the company's approach to modelling survival.

5.2.12 Company's approach to sensitivity analysis

The CS provides deterministic sensitivity analyses in the form of tornado diagrams, which examine the impact of raising and lowering individual parameters, and scenario analyses, which explore alternative data sources and model assumptions. The CS also provided a probabilistic sensitivity analysis (PSA) using the original model submitted by the company but results from the PSA were not provided for the revised model provided following the clarification request.

The ERG notes that the parameter ranges used to generate the tornado diagram were arbitrarily set at either ±10% or ±25% of the base-case value for all parameters except the discount rates (varied from 1% to 6%) and the time to discontinuation hazard (varied from 0.0065529 to 0.0083562). The ERG also notes that the original company model did not incorporate the parameter uncertainty associated with the regression used to predict utilities within their PSA. This was included in the revised model submitted following the clarification request (see response to clarification question B14)[1] but the method used to sample the regression coefficients using the variance-covariance matrix was not one familiar to the ERG. The ERG's preferred method for sampling regression coefficients, which are usually correlated, would be to assume that they follow a multivariate normal distribution which can be sampled using excel functions provided by the Centre for Bayesian Statistics in Health Economics (CHEBS).[50] The ERG compared the sampled utility values generated by the company's method with those generated using the CHEBs functions (for a fixed ICSI score). These are shown in Figure 13, where it can be seen that the distribution of utility values is much narrower when using the company's method. The ERG therefore concluded that uncertainty in the utility parameters is likely to have been underestimated in the company's PSA.

The ERG notes that in addition, several parameters have not been varied probabilistically within the PSA. These include the parameters for the survival functions used to estimate time to treatment discontinuation, the proportion of responders used to estimate median ICSI scores in responders and non-responders and the regression coefficients for the relationship between ICIS scores and resource use. The exclusion of these parameters from the PSA will also tend to underestimate the parameter uncertainty in the company's PSA.

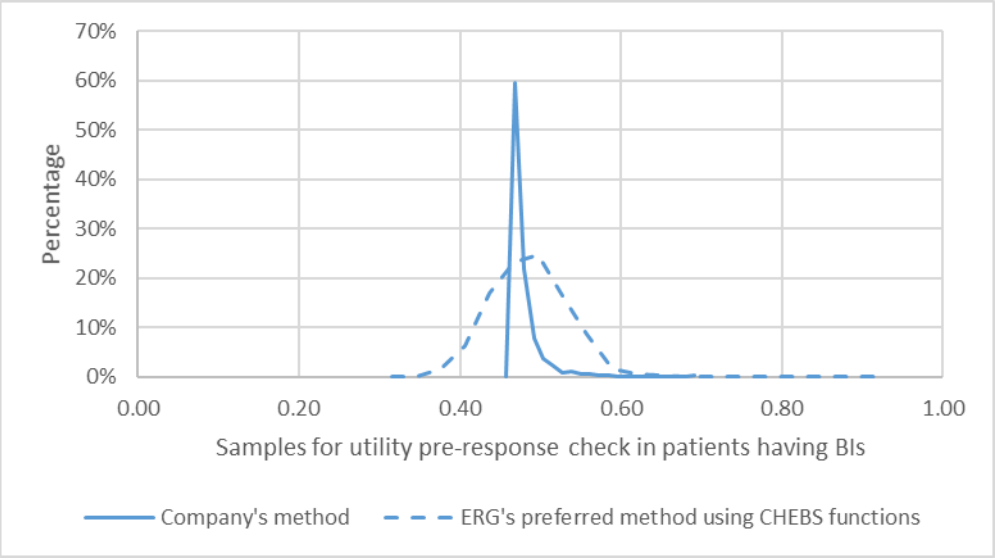


Figure 13: Comparison of utility values sampled as PSA inputs using the company method and the ERG's preferred method

The CS presents a scenario analysis in which a small percentage (2%) of those receiving BIs as subsequent treatment for a prolonged period go on to have a bladder procedure. This was implemented in the model by applying a fixed cost in the 10th year that patients receive second-line BIs. The cost of surgery was based on the weighted average of costs across eight different HRG codes. The scenario analysis did not adjust treatment-related or disease-related costs incurred following surgery but instead continued to apply the costs for BIs as second-line treatment until death or the model time horizon was reached. The scenario analysis also did not adjust QALYs to account for the impact of surgery on health outcomes.

The ERG considered that this scenario analysis lacked clinical face validity because it did not capture the impact of surgery on future utilities or costs and instead focused only on the one-off cost of the surgical procedure. However, it is unclear whether the ICER would increase or decrease if these factors were properly considered. Furthermore, the ERG did not understand why surgery was not an option for patients who have an inadequate response to BSC but who are unable to receive BIs. This would potentially bias the estimates of cost-effectiveness in favour of BSC under the company's current assumptions regarding surgery. However, the extent of any bias is likely to be small given that the company assumed that only 2% would go on to have surgical management. As the ERG's clinical advisors agreed that the frequency of surgical management for IC/BPS was low in current practice, the ERG did not conduct any exploratory analyses to explore the issues related to the modelling of surgical intervention in the scenario analysis as it was anticipated that the impact of any changes on the ICER would be small.

5.2.12 Cost effectiveness results

This section summarises the cost-effectiveness results presented in the CS. Following the clarification process, the company submitted a revised base case after rectifying a number of minor errors highlighted by the ERG. This section reports the updated base case results provided by the company.

Company's base-case analysis 1 (PPS versus bladder instillations)

Table 24 presents the estimates of cost-effectiveness generated using the company's revised model for the comparison of PPS versus bladder instillations. Compared to treatment with BIs, the probabilistic version of the model estimated that PPS would generate 0.25 additional QALYs at an additional cost of [redacted]; corresponding ICER of [redacted] per QALY gained. The deterministic model estimated a slightly lower ICER of [redacted] per QALY gained.

Table 24: Company's revised base-case results for PPS versus bladder instillations

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (per QALY gained)
Probabilistic model					
PPS	[redacted]	8.14	[redacted]	0.25	[redacted]
Bladder instillations	[redacted]	7.89	[redacted]	-	[redacted]
Deterministic model					
PPS	[redacted]	8.02	[redacted]	0.25	[redacted]
Bladder instillations	[redacted]	7.77	[redacted]	-	[redacted]

Company's base-case analysis 2 (PPS versus BSC):

For those patients who are not eligible for bladder instillations, due to them being inappropriate, poorly tolerated or unsuccessful, PPS is compared against best supportive care. Table 25 presents the base-case cost-effectiveness estimates of PPS versus BSC generated from the company's revised model. The probabilistic model estimates that PPS generates 0.33 additional QALYS in comparison to BSC, at an additional cost of [redacted]. This results in a much higher ICER than the PPS versus BI scenario of [redacted] per QALY gained. The deterministic version of the model estimated a slightly higher ICER of [redacted]

Table 25: Company's revised base-case results for PPS versus BSC

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (per QALY gained)
Probabilistic model					
PPS	[redacted]	8.45	[redacted]	0.33	[redacted]
BSC	[redacted]	8.12	[redacted]	-	[redacted]
Deterministic model					
PPS	[redacted]	8.34	[redacted]	0.32	[redacted]
BSC	[redacted]	8.02	[redacted]	-	[redacted]

5.2.13 Sensitivity analyses

The ERG notes that the following factors were significant drivers of cost-effectiveness based on the CS:

- Utilities in responders and non-responders
- Durability of response in responders to BSC (i.e. the time at which the treatment response recedes)

- Frequency of BI administration in the long-term
- Administration costs for BIs
- Inclusion of studies in broader BPS when estimating treatment effectiveness of PSS vs placebo
- Utilities and costs determined by ICPI instead of ICSI (particularly in PPS vs BSC)
- Application of baseline utilities and costs in non-responders
- Rates of self-administration of BIs
- Choice of time to treatment discontinuation curve (particularly in PPS vs BSC).

The company provided results of the deterministic and probabilistic sensitivity analysis within the original CS. However, updated versions of these were not provided following the company's update of their base-case model submitted with their response to the clarification letter. The PSA and DSA results presented in this section were generated by the ERG using the company's revised model.

Company's probabilistic sensitivity analysis (base-case analysis 1)

Figure 14 and Figure 15 show the cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) for PPS versus bladder instillations. The probability that PPS produces more net benefit than BIs at willingness to pay thresholds of £20,000 and £30,000 per QALY gained is 0.54 and 0.61, respectively. Although the ERG notes that these estimates should be interpreted with caution given their concern that the parameter uncertainty has underestimated in the PSA (see Section 5.2.12).

[redacted]

Figure 14: [redacted]

Figure 15:

Company's deterministic sensitivity analysis (base-case analysis 1)

The company presented the results of the DSA in the form of a tornado diagram, which was reproduced by the ERG using the company's revised model (Figure 16). Based on varying the parameters chosen in this analysis, the ICER is estimated to range from per QALY gained. The largest influences on the ICER were the utility of responders to both PPS and BI and administration costs of bladder instillations, both of which had a corresponding ICER above per QALY gained.

Figure 16:

Company's scenario analysis (base-case analysis 1)

The company conducted a number of scenario analyses in the original CS, which the ERG have updated using the revised model provided following the clarification process (see Table 26). The results of the scenario analyses suggest that the ICER is most sensitive to changes in parameters affecting the overall costs of bladder instillations, such as, the frequency of bladder administrations (post initial first month treatment) and bladder instillations being self-administered by the patient (implemented in the model through a reduction in the administration cost of bladder instillations). The ICER was also sensitive to using meta-analysed response rates for PPS that include two wider population trials; however, the populations included in these trials were outside of the NICE scope.

Table 26: Company scenario analysis for PPS versus bladder instillations, reproduced by the ERG using the company's revised model

Scenario	PPS costs	PPS QALYs	BI costs	BI QALYs	Incremental Cost	Incremental QALYs	ICER
Base-case		**8.021**	**£71,641**	**7.771**		**0.25**	
ICPI based utilities and background costs		8.365	£70,754	8.106		0.26	

Utilities from literature - (Cervigni 2017)[30]	[redacted]	5.611	£71,641	5.358	[redacted]	0.25	[redacted]
Lifetime horizon	[redacted]	11.653	£105,935	11.383	[redacted]	0.27	[redacted]
Discounting 1.5%	[redacted]	9.575	£85,542	9.292	[redacted]	0.28	[redacted]
Using least expensive product for BI (subsequent treatment)	[redacted]	8.021	£68,737	7.771	[redacted]	0.25	[redacted]
10% self-administration of BIs	[redacted]	8.021	£68,136	7.771	[redacted]	0.25	[redacted]
3-month response check	[redacted]	8.027	£71,588	7.788	[redacted]	0.24	[redacted]
Time horizon - 5 years	[redacted]	2.606	£22,812	2.474	[redacted]	0.13	[redacted]
Time horizon - 10 years	[redacted]	4.762	£41,698	4.565	[redacted]	0.20	[redacted]
Time horizon - 15 years	[redacted]	6.542	£57,882	6.310	[redacted]	0.23	[redacted]
Surgery as part of subsequent treatment	[redacted]	8.021	£71,675	7.771	[redacted]	0.25	[redacted]
Response rate for PPS including 2 wider population clinical trials	[redacted]	8.078	£70,929	7.896	[redacted]	0.18	[redacted]
Frequency BI administrations (post 1st month) set to 6 weeks (base-case is 4 weeks)	[redacted]	8.021	£69,149	7.771	[redacted]	0.25	[redacted]
Weibull distribution for time-to-discontinuation data	[redacted]	7.999	£71,838	7.761	[redacted]	0.24	[redacted]
Log-normal distribution for time-to discontinuation data	[redacted]	8.012	£71,504	7.768	[redacted]	0.24	[redacted]

Company's probabilistic sensitivity analysis (base-case analysis 2)

The company's revised model suggests that the probabilistic ICER for PPS versus BSC is [redacted] per QALY gained; this is lower than the company's deterministic ICER [redacted] per QALY gained). Figure 17 shows the cost-effectiveness plane for PPS versus BSC, with the corresponding CEAC shown

Figure 18. These show that for those patients unable to receive bladder instillations, the probability that PPS produces more net benefit than BSC at willingness to pay thresholds of £20,000 and £30,000 is 0.15 and 0.33, respectively. Although the ERG notes that these estimates should be interpreted with caution given their concern that the parameter uncertainty has underestimated in the PSA (see Section 5.2.12).

Figure 17: [redacted]

Figure 18:

Company's deterministic sensitivity analysis (base-case analysis 2)

The company's revised model DSA (Figure 19) found that the ICER was strongly impacted by changes in the utility of non-responders for both PPS and BSC with both higher and lower values, lowering the ICER significantly. However, changes in the utility of responders to PPS to the lower value used in the DSA resulted in much higher ICER).

Figure 19:

Company's scenario analysis (base-case analysis 2)

Table 27 shows the results of the scenario analysis for PPS versus BSC, reproduced by the ERG using the company's revised model. The results of the scenario analyses suggest that the ICER is particularly sensitive to changes in the duration of the receding effect for the placebo response of BSC (ICER ranging from [redacted] per QALY gained). Utilities used in the model also had a large impact on the ICER, with utilities based on Cervigni 2017[30] largely reducing the ICER to [redacted] per QALY whilst basing utilities and background costs on ICPI scores as opposed to ICSI reduced the ICER to [redacted] per QALY gained.

Table 27: Company scenario analysis for PPS versus BSC, reproduced by the ERG using the company's revised model

Scenario	PPS costs	PPS QALYs	BSC costs	BSC QALYs	Incremental Cost	Incremental QALYs	ICER
Base-case	[redacted]	8.337	£23,448	8.017	[redacted]	0.32	[redacted]
ICPI based utilities and background costs	[redacted]	8.470	£23,017	8.030	[redacted]	0.44	[redacted]
Utilities from literature - (Cervigni 2017)[30]	[redacted]	4.600	£23,448	3.647	[redacted]	0.95	[redacted]
Lifetime horizon	[redacted]	12.151	£34,487	11.802	[redacted]	0.35	[redacted]
Discounting 1.5%	[redacted]	9.959	£28,039	9.592	[redacted]	0.37	[redacted]
BSC effect receding at 6 months	[redacted]	8.337	£23,501	8.007	[redacted]	0.33	[redacted]
BSC effect receding at 5 years	[redacted]	8.337	£23,055	8.097	[redacted]	0.24	[redacted]
BSC effect not receding	[redacted]	8.337	£22,344	8.167	[redacted]	0.17	[redacted]
Baseline utility and background costs given to non-responders	[redacted]	7.988	£26,368	7.595	[redacted]	0.39	[redacted]
3-month response assessment	[redacted]	8.349	£23,395	8.025	[redacted]	0.32	[redacted]
Time horizon - 5 years	[redacted]	2.687	£7,484	2.544	[redacted]	0.14	[redacted]
Time horizon - 10 years	[redacted]	4.929	£13,742	4.690	[redacted]	0.24	[redacted]
Time horizon - 15 years	[redacted]	6.788	£19,012	6.496	[redacted]	0.29	[redacted]
Response rate including 2 wider population clinical trials	[redacted]	8.316	£23,436	8.020	[redacted]	0.30	[redacted]
Weibull distribution for time-to-discontinuation data	[redacted]	8.319	£23,448	8.017	[redacted]	0.30	[redacted]
Log-normal distribution for time-to discontinuation data	[redacted]	8.330	£23,448	8.017	[redacted]	0.31	[redacted]
Surgery – same as base-case as not affected							

5.2.14 Model validation and face validity check

The ERG validated the implementation of the sampling of time to discontinuation by plotting the cumulative survival functions from the samples generated by the VBA code. In doing so, it was identified that the survival function for the Weibull distribution was not correctly implemented in the company's original base-case. This was due to an incorrect translation between two different parameterisations of the Weibull survival function. However, the company corrected this in their model submitted with their response to the clarification request,[1] and so this error does not affect the results presented in Sections 5.2.12 and 5.2.13.

The ERG validated the VBA code by stepping through the code for patients with different trajectories (i.e. responders and non-responders in each arm), by using the locals window to observe the changes to the costs and QALYs at each event and by checking lifetime costs and QALYs for selected individual patients using the patient-level model output data. In doing so, it was identified that the time to response check was being converted from months to years for all instances where it was being used in the VBA except when calculating the "other costs" accrued between the time of the response check and the time of discontinuation. This resulted in some cases in the cost per annum being multiplied by a negative period of time. The company corrected this for the VBA code used to run the base-case analysis in their model submitted with their response to the clarification request,[1] and so this error does not affect the results presented in Sections 5.2.12 and 5.2.13. However, the correction was not carried through to the separate VBA subroutines used to run the scenario analysis and therefore these were corrected by the ERG (see Appendix 2 for details of the correction).

The ERG checked the patient-level results against the expected values for individual patients with various trajectories based on the ERG's understanding of the CS.[1] The ERG was satisfied that the model was behaving in the expected manner at the individual level.

The ERG rebuilt the model as a state-transitions model to determine whether this was feasible without altering the conceptual model and to provide an external validation of the company's DES approach. The ERG was satisfied that the ICERs were sufficiently close to exclude there being a significant unidentified error in the DES or a significant error in the ERG's understanding of the conceptual model. The ERG notes that it was possible to rebuild the model as a simple state-transition model without the need to include any non-Markovian fixes or time-dependent transition probabilities when using the exponential time to treatment discontinuation curve. However, implementation of alternative parametric forms with time varying risks of discontinuation, such as the log normal, would have required the use of time-dependent transition probabilities. The ERG considers that a state-transition

approach would have been more parsimonious but that this does not mean that the DES approach is incorrect.

The DES model did not allow for the reporting of costs and QALYs according to the individual's trajectory through the model, which would be analogous to the costs and QALYs accrued in various health states for a state transition model. To address this, the ERG identified the proportion of the QALY gain associated with additional patients who respond in PPS vs BSC by examining patient-level QALY gains. It found that only 53% of the QALY gains for PPS versus BI and 53% of the QALY gains for PPS versus BSC were accrued due to the higher rate of response achieved by PPS. In the comparison against BI, the remainder of the QALY gains were associated with the utility decrement for "previous BI usage" from the regression analysis of the patient survey data.[1] In the comparison against BSC, the remainder of the QALY gains were related to the assumption that responders to BSC benefit for a maximum of 12 months whereas responders to PPS benefit until they discontinue.

To check the internal validity of the model, the ERG calculated the proportion of responders from the patient-level results and noted that the average rate of responders was 33.8%, 22.8% and 16.5% based on the first 10,000 patients sampled whereas the input values for these parameters were 33.1%, 22.0% and 15.8% respectively. The ERG suspected that this slightly discrepancy was due to the stochastic nature of the model whereby stable outputs are only achieved if sufficient patients have been simulated. The ERG conducted a large run of 100,000 patients and found that the ICERs based on the first 10,000 patients was within £500 per QALY of the ICER based on the larger run of 100,000 patients. The ERG was therefore satisfied that the results provided by the model, which were based on 10,000 patients, were sufficiently accurate for decision making.

The ERG noticed that there were a number of minor discrepancies between the values provided in the CS[1] and those included in the model (e.g. ICSI scores in Table 41, mean and standard deviation for time to death in Table 57), but the correct values had been included in the model. The ERG also noticed a minor discrepancy between the source study and the values used in the CS[1] for the mean starting age based on the data from Sant *et al.* 2003,[17] but the difference was too small to make any difference to the model (45.57 years vs 45.41 years with the life-expectancy data being based on patients aged 45 years).

5.3 Exploratory and sensitivity analyses undertaken by the ERG

5.3.1 ERG's exploratory analysis- methods

Following concerns highlighted in Section 5.2, the ERG undertook seven sets of exploratory analyses by implementing changes to the company's revised model. Two of these changes were not applicable

to the comparison of PPS against BIs because they related to the modelling of BSC and one of these changes was not applicable to the comparison of PPS against BSC because it related to the modelling of BI. Combining all of the changes applicable to each comparison forms the ERGs preferred base-case for that comparison. The seven changes are discussed in turn below.

Exploratory analysis 1: Use of all discontinuations from Hanno et al (1997)[27] *for survival analysis of time to treatment discontinuation*

As noted in Section 5.2.10, the ERG had concerns with how the time to discontinuation data provided by Hanno *et al.*[27] had been interpreted by the company when estimating the cumulative probability of remaining on treatment. The ERG therefore conducted a scenario analysis which incorporates their preferred interpretation of the data which include: using all patients know to have discontinued from the Hanno *et al.* (1997) study;[27] censor patients at 60 months for those reported to have follow up for 60+ months and time measured from 6 months. In exploratory analysis 1, the ERG used the company's preferred parametric function which was the exponential. The ERG's preferred parametric function is considered in exploratory analysis 7.

Exploratory analysis 2: Switch to 6-weekly dosing for first-line BIs after first year of treatment and 6-weekly for all 2nd line BIs

The ERG believes that long-term dosing of BIs is likely to be overestimated in the company's model and that dosing with BIs will decrease in the long-term (Section 5.2.9). The ERG therefore implemented a switch to 6-weekly dosing from year 1 for first-line BI treatment and 6-weekly BI treatments for all those receiving BI as subsequent treatment.

Exploratory analysis 3: Use regression for utility based on ICSI scores which excludes term for prior usage of BI

As noted in Section 5.2.9, the ERG believes that applying a utility decrement to patients who responded to having current/recent treatment with bladder instillation is inappropriate, as the differences detected through the survey may reflect differences in patient characteristics as opposed to differences associated directly with treatment with bladder instillations. In additional, it was unclear if there is any decrement associated with PPS treatment and the ERG was not satisfied with how missing data for recent treatment with bladder instillations in the survey had been handled. Therefore, the ERG applied the company's alternative regression which included coefficients for age and ICSI, but no coefficient for bladder instillations (model identified as "Twopm 1" in Table 86 of the CS).

Exploratory analysis 4: Use of a lifetime horizon

As the survival function predicts that some patients will remain on treatment at the end of the 20-year time horizon, the ERG used a lifetime horizon to ensure all costs and benefits associated with treatment are captured within the model.

Exploratory analysis 5: Return to baseline utilities and costs for non-responders when BSC is second-line option (PPS versus BSC scenario only)

The ERG does not believe that it is clinically valid to assume that patients not responding to BSC or PPS would benefit from an improvement in ICSI scores, and an associated improvement in both cost and utilities, for the remainder of their lifetime. Therefore, the ERG instead assumed that patients return to baseline utility and cost levels following no response to BSC.

Exploratory analysis 6: Switch off receding baseline response for BSC (PPS versus BSC scenario only as already implemented in PPS versus BI base-case)

As noted in Section 5.2.7, the ERG believe it is inconsistent to assume that the placebo response for BSC would cease at 12 months yet responses for BI and PPS remain durable for the remainder of treatment. The ERG believed a more consistent approach would be to apply the same durability of response for all arms and therefore removed the receding baseline response for BSC.

Exploratory analysis 7: Use of log-normal function to model time to discontinuation

Based on the results of statistical fit (AIC and BIC), time to treatment discontinuation was modelled using a log-normal survival function for the dataset based on the ERG's preferred interpretation of the time to discontinuation data presented by Hanno *et al.*[27] (see exploratory analysis 1).

5.3.2 Results of ERG's exploratory analysis

All of the results presented below have been generated using mean parameter inputs (i.e. using the 'deterministic' rather than the PSA version of the model).

ERG's preferred analysis 1 (PPS versus bladder instillations)

Results for PPS versus BIs are presented in Table 28 as individual changes to the company's revised model, with all changes (1 to 7) then combined to give the ERGs preferred base-case.

Table 28: Results of ERG's preferred analysis for PPS versus BI

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
Company's base-case (revised base-case model, deterministic)					
PPS	[redacted]	8.021	[redacted]	0.25	[redacted]
Bladder instillations	[redacted]	7.771	[redacted]	-	[redacted]
ERG exploratory analysis 1: Use of all discontinuation data for survival analysis inputs					
PPS	[redacted]	7.797	[redacted]	0.13	[redacted]
Bladder instillations	[redacted]	7.671	[redacted]	-	[redacted]
ERG exploratory analysis 2: Switch to 6 weekly dosing for first line BIs after first year of treatment and 6 weekly for all 2nd line BIs					
PPS	[redacted]	8.021	[redacted]	0.25	[redacted]
Bladder instillations	[redacted]	7.771	[redacted]	-	[redacted]
ERG exploratory analysis 3: Utility regression used excludes 'had BI' coefficient					
PPS	[redacted]	7.899	[redacted]	0.09	[redacted]
Bladder instillations	[redacted]	7.809	[redacted]	-	[redacted]
ERG exploratory analysis 4: Lifetime horizon					
PPS	[redacted]	11.653	[redacted]	0.27	[redacted]
Bladder instillations	[redacted]	11.383	[redacted]	-	[redacted]
ERG exploratory analysis 5: Return to baseline utilities and costs for non-responders when BSC is second line option (*Not applicable)*					
ERG exploratory analysis 6: Switch off receding baseline response for BSC (*Not applicable)*					
ERG exploratory analysis 7: Log-normal distribution for time to discontinuation					
PPS	[redacted]	7.811	[redacted]	0.13	[redacted]
Bladder instillations	[redacted]	7.679	[redacted]	-	[redacted]
ERG's preferred base-case (including all ERG individual amendments 1-7)					
PPS	[redacted]	11.373	[redacted]	0.04	[redacted]
Bladder instillations	[redacted]	11.331	[redacted]	-	[redacted]

The updates made to the time to discontinuation data (log-normal distribution and ERG preferred survival analysis inputs) do not have a large impact on the ICER, but both result in a more favourable ICER relative to the company's revised base-case. The application of a lifetime horizon resulted in a slight increase in the ICER; however, the largest increases resulted from individual changes to bladder instillation dosing (exploratory analysis 2, ICER [redacted] per QALY gained) and the application of the utility regression excluding bladder instillation use (exploratory analysis 3, ICER [redacted] per QALY gained). The ERG's preferred base-case, which combines all individual changes (exploratory analyses 1-7), resulted in an ICER of [redacted] per QALY gained; this is significantly higher than the company's revised base-case of [redacted] per QALY gained.

ERG's preferred analysis 2 (PPS versus BSC)

Results for the ERG's exploratory analysis for PPS versus BSC are shown in Table 29, presented as individual changes to the company's revised model, with all changes then combined to give the ERGs preferred base-case.

Table 29: Results of ERG's preferred analysis for PPS versus BSC

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
Company's base-case (revised base-case model, deterministic)					
PPS	[redacted]	8.337	[redacted]	0.32	[redacted]
BSC	[redacted]	8.017	[redacted]	-	[redacted]
ERG exploratory analysis 1: Use of all discontinuation data for survival analysis inputs					
PPS	[redacted]	8.151	[redacted]	0.13	[redacted]
BSC	[redacted]	8.017	[redacted]	-	[redacted]
ERG exploratory analysis 2: Switch to 6 weekly dosing for first line BIs after first year of treatment and 6 weekly for all 2nd line BIs (*not applicable)*					
ERG exploratory analysis 3: Utility regression used excludes 'had BI' coefficient					
PPS	[redacted]	7.540	[redacted]	0.34	[redacted]
BSC	[redacted]	7.202	[redacted]	-	[redacted]
ERG exploratory analysis 4: Lifetime horizon					
PPS	[redacted]	12.151	[redacted]	0.35	[redacted]
BSC	[redacted]	11.802	[redacted]	-	[redacted]
ERG exploratory analysis 5: Non-responders receiving BSC return to baseline utility and cost values					
PPS	[redacted]	7.988	[redacted]	0.39	[redacted]
BSC	[redacted]	7.595	[redacted]	-	[redacted]
ERG exploratory analysis 6: Receding effect of placebo response switched off					
PPS	[redacted]	8.337	[redacted]	0.17	[redacted]
BSC	[redacted]	8.167	[redacted]	-	[redacted]
ERG exploratory analysis 7: Log-normal distribution for time to discontinuation					
PPS	[redacted]	8.163	[redacted]	0.15	[redacted]
BSC	[redacted]	8.017	[redacted]	-	[redacted]
ERG's preferred base-case (including all ERG individual amendments 1-7)					
PPS	[redacted]	10.129	[redacted]	0.11	[redacted]
BSC	[redacted]	10.022	[redacted]	-	[redacted]

The use of the ERG's preferred survival analysis inputs increased the ICER from a base-case of [redacted] per QALY gained. This was a much greater impact than observed in the PPS vs BI comparison because in this scenario only the PPS arm is altered by the time to discontinuation data. When using the ERG's preferred interpretation of the data from Hanno *et al.*,[27] the switch from the exponential to the log-normal parametric function had a small impact and decreased the ICER slightly relative to the company's preferred choice of the exponential function (scenario 7 compared to scenario 1). Non-responders returning to baseline utility values (exploratory analysis 5) resulted in a decrease in the ICER to [redacted] per QALY gained. The assumption regarding the receding effect of the placebo response for BSC was shown to be the key driver of the ICER (exploratory analysis 6).

The ERG's preferred base-case combining all scenarios (1-7) results in a substantially higher ICER of [redacted] per QALY gained compared to the company's revised base-case analysis.

5.3.3 Additional sensitivity analysis undertaken using the ERG's preferred base-case model

Additional sensitivity analyses were also undertaken using the ERG's preferred base-case model in order to explore different assumptions made within the model:

- All costs based on baseline ISCI scores removing the relationship between response to treatment and costs
- Explore different baseline response rates through changes to response rate of BSC based on upper and lower confidence intervals reported in the literature.
- Urology outpatient cost used for administration of bladder instillations (PPS versus BI only)
- 80% of patients self-administer bladder instillations (PPS versus BI only)

Again, all of the results presented below have been generated using mean parameter inputs (i.e. using the 'deterministic' rather than the PSA version of the model).

Additional sensitivity analysis results: PPS versus BI

Table 30: Additional sensitivity analysis undertaken using ERG preferred base-case model for PPS versus BI

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
ERG preferred base-case					
PPS	[redacted]	11.373	[redacted]	0.04	[redacted]
BI	[redacted]	11.331	[redacted]	-	[redacted]
ERG sensitivity analysis 1: All costs based on baseline ISCI scores removing the relationship between response to treatment and costs					
PPS	[redacted]	11.373	[redacted]	0.04	[redacted]
BI	[redacted]	11.331	[redacted]	-	[redacted]
ERG sensitivity analysis 2: Response rate of BSC set to 5% (equal to lower confidence interval from Mulholland *et al.* 1990)[14]					
PPS	[redacted]	10.847	[redacted]	0.02	[redacted]
BI	[redacted]	10.829	[redacted]	-	[redacted]
ERG sensitivity analysis 3: Response rate of BSC set to 32% (equal to upper confidence interval from Parsons *et al.* 1987)[16]					
PPS	[redacted]	12.095	[redacted]	0.06	[redacted]
BI	[redacted]	12.031	[redacted]	-	[redacted]
ERG sensitivity analysis 4: Urology outpatient cost used for administration of bladder instillations					
PPS	[redacted]	11.373	[redacted]	0.04	[redacted]
BI	[redacted]	11.331	[redacted]	-	[redacted]
ERG sensitivity analysis 5: 80% rate of self-administration of bladder instillations					
PPS	[redacted]	11.373	[redacted]	0.04	[redacted]
BI	[redacted]	11.331	[redacted]	-	[redacted]

Table 30 presents the results for the ERG's additional sensitivity analysis conducted using the ERG's preferred base-case model. Removing the assumption regarding a relationship between patients' response to treatment and healthcare costs (sensitivity analysis 1) leads to a small increase in the ICER, as the resulting increase in costs for PPS is slightly larger than that in BIs.

The ERG explored its concerns with the reliability of the data used for the response rate of BSC, as previously mention in Section 5.2.6, through conducting sensitivity analyses on the percentage of BSC responders used in the model (sensitivity analysis 2 and 3). Upper and lower extremes were used based on the highest and lowest confidence intervals reported in the literature used to form the meta-analysis. The ICER was very sensitive to changes in the response rate of BSC, but the direction of change is somewhat counterintuitive. It can be seen that the lower response rate results in a smaller QALY gain, as the difference in the absolute number of responders between PPS and BIs decreases. However, the ICER reduces because the incremental costs decrease more than the incremental QALYs giving an ICER of [redacted] per QALY gained. This is because patients on PPS have a reduction in costs when they fail to respond, but patients on BSC have a slight increase in costs when they fail to respond (see Figure 2). The opposite is true in sensitivity analysis 3, with the increase in response rate for BSC resulting in a larger QALY gain but a higher ICER of [redacted] per QALY gained. Table 30 also shows that the ICER was also sensitive to changes in administration of bladder instillations (sensitivity analyses 4 and 5). Using a urology specific outpatient cost (£151) for administration of bladder instillations increased the ICER to [redacted] per QALY gained, compared to [redacted] per QALY gained when the company's cost of £183 was used. Given that the ERG's clinical experts reported varying experiences regarding the proportion of patients who self-catheterise for BIs, ranging from none to 80%, the ERG explored a scenario in which a high proportion of patients self-administer BIs, resulting in a large increase in the ICER.

Additional sensitivity analysis results: PPS versus BSC

Table 31: Additional sensitivity analysis undertaken using ERG preferred base-case model for PPS versus BSC

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
ERG preferred base-case					
PPS	[redacted]	10.129	[redacted]	0.11	[redacted]
BSC	[redacted]	10.022	[redacted]	-	[redacted]
ERG sensitivity analysis 1: All costs based on baseline ISCI scores removing the relationship between response to treatment and costs					
PPS	[redacted]	10.129	[redacted]	0.11	[redacted]
BSC	[redacted]	10.022	[redacted]	-	[redacted]
ERG sensitivity analysis 2: Response rate of BSC set to 5% (equal to lower confidence interval from Mulholland *et al.* 1990)[14]					
PPS	[redacted]	9.987	[redacted]	0.04	[redacted]
BSC	[redacted]	9.951	[redacted]	-	[redacted]
ERG sensitivity analysis 3: Response rate of BSC set to 32% (equal to upper confidence interval from Parsons *et al.* 1987)[16]					
PPS	[redacted]	10.344	[redacted]	0.22	[redacted]
BSC	[redacted]	10.122	[redacted]	-	[redacted]

Table 31 reports the results of additional sensitivity analyses for PPS versus BSC using the ERG's preferred base-case analysis. As in the PPS versus BI analysis, the ICER was somewhat sensitive to changes in the assumptions on the relationship between treatment response and healthcare costs, resulting in a marginally higher ICER. Implementing changes to the response rate of BSC again had a large impact on the ICER, with a lower response rate of 5% resulting in a higher ICER of [redacted] per QALY gained and a higher response rate of 32% for BSC resulting in a lower ICER of [redacted] per QALY gained. The ERG notes that the impact on the ICER for PPS versus BSC is in opposite direction to that observed in the PPS versus BI scenario. This is because in the comparison of PPS versus BSC, the lower response rate results in a larger proportionate reduction in incremental QALYs than incremental costs.

5.4 Conclusions of the cost effectiveness section

The ERG was satisfied that the only cost-effectiveness paper identified in the company's review of published cost-effectiveness analyses was not sufficiently applicable to the decision problem specified in the scope and therefore a *de novo* analysis was necessary. The ERG had some concerns regarding whether the company's review of studies reporting costs, resource use and HRQoL data had been adequate.

The ERG considered that the company's *de novo* analysis was relevant to decision problem specified in the final NICE scope for this appraisal in terms of the population considered and the interventions and comparators considered.

The ERG was broadly satisfied with the structure of the company's economic model, although it considered that the use of a DES structure was unnecessary in this case and a more parsimonious model could have been constructed using a state transition modelling approach.

The ERG identified several important uncertainties in the model inputs which have the potential to have a large impact on the ICER. The key areas of concern identified by the ERG were:

- The application of a utility decrement for patients receiving BIs, estimated from the patient survey, which the ERG did not consider was robust given that the handling of missing data on BI usage had not been adequately explored in the analysis of the survey data.
- Uncertainty surrounding the likely rate of response in patients receiving BSC in clinical practice which affects the absolute difference in response attributable to PPS in the model
- Inconsistent assumptions around the durability of response in those receiving BSC and those receiving either PPS or BIs
- The assumption that 4 weekly administrations of BIs continues indefinitely when the ERG believes that the frequency of administration is likely to fall over time
- The underestimation of treatment discontinuation rates which affects the lifetime treatment costs, particularly for PPS versus BSC
- The assumption that patients who do not respond to BSC have some long-term persistent utility gain relative to baseline
- Low rates of self-administration for BIs which may overestimate costs relative to established clinical practice in some parts of the NHS

The impact on the ICER of these concerns was demonstrated in the ERG's exploratory analysis which produced an ERG preferred ICER of [redacted] per QALY for PPS vs BSC and [redacted] per QALY for PPS vs BI. However, there were some data inputs and assumptions that remain uncertain and which the ERG explored in further sensitivity analyses. These were found to have the potential to increase the ICER to [redacted] for PPS versus BI and up to [redacted] for PPS versus BSC.

The ERG also had additional concerns regarding the robustness of the data used to inform the model which were related to:

- the use of an unadjusted indirect comparison between PPS and BI to determine relative response rate

- the use of data from the broader population with BPS rather than the population with IC/BPS to estimate the efficacy of BI versus placebo
- The assumption that the long-term cumulative rate of response to second-line BIs is equivalent to the short-term response to first-line BIs
- the assumption of equal discontinuation rates for PPS and BI
- the method used to estimate ICSI scores for responders and non-responders
- the choice of HRG costs applied in the patient survey and the robustness of the relationship between ICSI score and disease-related costs
- the under estimation of parameter uncertainty within the PSA

6 END OF LIFE

The end of life criteria are not considered relevant in this appraisal as the company has not made a case that they should be considered and the ERG is not aware of any evidence that IC/BPS has any impact on life expectancy.

7 OVERALL CONCLUSIONS

The company's systematic review of clinical effectiveness suggests PPS to be significantly better than placebo for treating IC/BPS on improvement global response assessment in some RCTs but not others. Similar results in favour of PPS were also evident for non-VAS pain.

The company's systematic review of clinical effectiveness also indicated there to be no statistically significant between-group differences in: mean O'Leary-Sant Interstitial Cystitis Symptom Index and Problem Index scores (*p*-values or CI, not reported), mean daily urinary frequency (*p*-value not reported and $p=0.06$, CIs not reported), mean urinary volume and void outcomes (*p*-values or CIs, not reported), or mean nocturia (*p*-values or CIs, not reported), reported by the RCTs of PPS compared to placebo for treating IC/BPS.

The ERG's critique of the clinical effectiveness evidence identified that study quality in one of the four included RCTs of PPS for IC/BPS was unclear regarding: allocation concealment, details of who was blind, and numbers of patients withdrawing from treatment groups. As such, the ERG considers that the results from this RCT should be interpreted with caution.

The ERG notes that there is some author commonality across all four RCTs of PPS for IC/BPS and that no further published studies, undertaken by an independent study group, have attempted to validate the results of the four RCTs of PPS for IC/BPS.

The ERG also notes limitations in the reporting of outcome data in the PPS RCT trial reports (no interval estimates and *p*-values for non-significance often not reported).

The company's pairwise meta-analysis across RCTs suggests PPS to be significantly better than placebo for treating IC/BPS on improvement in global response assessment (RR, 2.09; 95%CI, 1.47 to 2.97; fixed effect).

The ERG has some concerns with the pairwise meta-analyses that were performed by the company (choice of scale for the analysis, the use of hypothesis testing to assess heterogeneity, and the use of a fixed effect model in the absence of evidence that there is not between study heterogeneity). The company also undertook an indirect comparison between PPS and Uracyst using the Bucher method with the placebos as the reference treatment. This gave an effective RR for PPS versus BI of 1.50. The ERG accepts the arguments suggested by the company for not performing an NMA. However, the ERG does not believe that the Bucher approach mitigates all of the concerns associated with performing an NMA, including: not using a single model to incorporate random effects; making the

assumption of asymptotic normality when making inferences and characterising uncertainty about the relative treatment effect used in the economic model.

An NMA or ITC of AEs was not undertaken by the company. Instead, a summary of AEs reported in the four RCTs of PPS for IC/BPS were presented. The ERG notes that >50% of patients in both PPS and PBO treatment groups were reported as experiencing moderate AEs in one RCT. However, the ERG accepts the company's conclusion that PPS is well tolerated, given that clinical advice received by the ERG is that AEs with PPS are rare.

The ERG considered the company's economic model to be consistent with the decision problem specified in the NICE scope[2]. The company's model is generally in line with the NICE reference case,[38] although the ERG had some concerns regarding the efficacy data used to inform the model. These included the use of data from trials in the broader BPS population to estimate the efficacy of BIs versus placebo and the methods used for the ITC.

The ERG considered that the structure of the company's model was appropriate. However, the ERG had concerns regarding some of the data inputs and assumptions used in the model. Several areas of uncertainty were identified which have the potential to have a significant impact on the ICER. These included: the likely response rate for BSC in clinical practice; the durability of response in those who have responded to BSC at 6 months; the expected ICSI scores for responders and non-responders and in particular the expected ICSI scores in patients who do not respond to BSC; the rate of persistence with treatment in the long-term for both PPS and BIs; the frequency of treatment with BIs in the long-term; the setting for administering BIs (outpatient versus day-case versus self-administered at home) and whether there is a utility decrement associated with treatment with BIs.

Based on the ERG's exploratory analyses which examined many of these factors, the ERG considers that the ICERs are likely to be much higher than presented in the company's base-case analysis and that there remains substantial uncertainty around the cost-effectiveness of PPS for treating IC/BPS.

7.1 Implications for research

The ERG believes that a three arm open-label study comparing PPS, BIs and placebo would provide valuable additional evidence. This would address the fact that no prospective Phase 3 study of PPS was conducted following the RCT by Sant *et al*. It would also address the lack of any direct evidence comparing PPS to BIs which are the current standard of care for IC/BPS. Any such trial should aim to collect evidence on the patient's global response to treatment (GRA) and the impact of treatment on HRQoL using both generic and disease specific instruments.

8 REFERENCES

1. Consilient Health. Pentosan polysulfate sodium for treating bladder pain syndrome [ID1364]; 2019.
2. National Institute for Health and Care Excellence. Pentosan polysulfate sodium for treating bladder pain syndrome - Final Scope. https://www.nice.org.uk/guidance/gid-ta10335/documents/final-scope [Accessed 25 January 2019] (Accessed
3. Engeler D, Baranowski AP, Borovicka J, Cottrell AM, Dinis-Oliveira P, Elneil S, *et al.* Chronic Pelvic Pain EAU Guidelines on. In; 2018.
4. Hanno PM. Interstitial cystitis-epidemiology, diagnostic criteria, clinical markers. *Reviews in urology* 2002;4 Suppl 1:S3-8.
5. van de Merwe JP, Nordling J, Bouchelouche P, Bouchelouche K, Cervigni M, Daha LK, *et al.* Diagnostic Criteria, Classification, and Nomenclature for Painful Bladder Syndrome/Interstitial Cystitis: An ESSIC Proposal. *European Urology* 2008;53:60-7.
6. Uddin M, Azhar F, Pearce I. Bladder pain syndrome. *Obstetrics, Gynaecology & Reproductive Medicine* 2018;28:333-8.
7. The Urology Foundation. Bladder-related Statistics. https://www.theurologyfoundation.org/professionals/healthcare-resources-and-reports/urology-resources/facts-and-figures/bladder-related-statistics [Accessed 25 January 2019] (Accessed
8. European Medicines Agency. Elmiron Summary of Product Characteristics. In; 2017.
9. Bade JJ, Rijcken B, Mensink HJ. Interstitial cystitis in The Netherlands: prevalence, diagnostic criteria and therapeutic preferences. *J Urol* 1995;154:2035-7; discussion 7-8.
10. Leppilahti M, Tammela TL, Huhtala H, Auvinen A. Prevalence of symptoms related to interstitial cystitis in women: a population based study in Finland. *J Urol* 2002;168:139-43.
11. Temml C, Wehrberger C, Riedl C, Ponholzer A, Marszalek M, Madersbacher S. Prevalence and correlates for interstitial cystitis symptoms in women participating in a health screening project. *Eur Urol* 2007;51:803-8; discussion 9.
12. Nickel JC. Interstitial cystitis: characterization and management of an enigmatic urologic syndrome. *Reviews in urology* 2002;4:112-21.
13. Warren JW, Brown J, Tracy JK, Langenberg P, Wesselmann U, Greenberg P. Evidence-based criteria for pain of interstitial cystitis/painful bladder syndrome in women. *Urology* 2008;71:444-8.
14. Mulholland SG, Sant GR, Hanno P, Staskin DR, Parsons L. Pentosan polysulfate sodium for therapyof interstitial cystitis. *Urology* 1990;35:552-8.
15. Parsons CL, Benson G, Childs SJ, Hanno P, Sant GR, Webster G. A Quantitatively Controlled Method to Study Prospectively Interstitial Cystitis and Demonstrate the Efficacy of Pentosanpolysulfate. *The Journal of Urology* 1993;150:845-8.
16. Parsons CL, Mulholland SG. Successful Therapy of Interstitial Cystitis with Pentosanpolysulfate. *The Journal of Urology* 1987;138:513-6.
17. Sant GR, Propert KJ, Hanno PM, Burks D, Culkin D, Diokno AC, *et al.* A Pilot Clinical Trial of Oral Pentosan Polysulfate And Oral Hydroxyzine in Patients With Interstitial Cystitis. *The Journal of Urology* 2003;170:810-5.
18. Holm-Bentzen M, Jacobsen F, Nerstrøm B, Lose G, Kristensen JK, Pedersen RH, *et al.* A prospective double-blind clinically controlled multicenter trial of sodium pentosanpolysulfate in the treatment of interstitial cystitis and related painful bladder disease. *The Journal of urology* 1987;138:503-7.
19. Nickel JC, Herschorn S, Whitmore KE, Forrest JB, Hu P, Friedman AJ, *et al.* Pentosan polysulfate sodium for treatment of interstitial cystitis/bladder pain syndrome: Insights from a randomized, double-blind, placebo controlled study. *Journal of Urology* 2015;193:857-62.
20. Nickel JC, Egerdie RB, Steinhoff G, Palmer B, Hanno P. A Multicenter, Randomized, Double-blind, Parallel Group Pilot Evaluation of the Efficacy and Safety of Intravesical Sodium Chondroitin Sulfate Versus Vehicle Control in Patients With Interstitial Cystitis/Painful Bladder Syndrome. *Urology* 2010;76:804-9.

21. Nickel JC, Hanno P, Kumar K, Thomas H. Second Multicenter, Randomized, Double-blind, Parallel-group Evaluation of Effectiveness and Safety of Intravesical Sodium Chondroitin Sulfate Compared With Inactive Vehicle Control in Subjects With Interstitial Cystitis/Bladder Pain Syndrome. *Urology* 2012;79:1220-5.
22. O'Leary MP, Sant GR, Fowler FJ, Whitmore KE, Spolarich-Kroll J. The interstitial cystitis symptom index and problem index. *Urology* 1997;49:58-63.
23. NICE. The guidelines manual: appendices B–I. 2012. https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-pdf-3304416006853 [Accessed 18 February 2019] (Accessed
24. Gillenwater JY, Wein AJ. Summary of the National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases Workshop on Interstitial Cystitis. Paper presented at: National Institutes of Health, Bethesda, Maryland, USA. August 28–29, 1987. .
25. Food and Drug Administration. A statistical and medical review of an amendment to the New Drug Application for Elmiron (Pentosan Polysulfate) NDA# 20193. *Journal* 1995. https://www.accessdata.fda.gov/drugsatfda_docs/nda/pre96/020193Orig1s000rev.pdf [Accessed 22 February 2019]
26. Committee for Medicinal Products for Human Use. elmiron Assessment Report. In. London; 2016.
27. Hanno PM. Analysis of long-term elmiron therapy for interstitial cystitis. *Urology* 1997;49:93-9.
28. Jepsen JV, Sall M, Rhodes PR, Schmidt D, Messing E, Bruskewitz RC. Long-term experience with pentosanpolysulfate in interstitial cystitis. *Urology* 1998;51:381-7.
29. Tutolo MA, E.; Castagna, G.; Klockaerts, K.; Plancke, H.; Ost, D.; Van der Aa, F.; De Ridder, D. A prospective randomized controlled multicentre trial comparing intravesical DMSO and chondroitin sulphate 2% for painful bladder syndrome/interstitial cystitis. *International Braz J Urol* 2017;43:134-41.
30. Cervigni MS, M.; Tenaglia, R.; Porru, D.; Ostardo, E.; Giammo, A.; Trevisan, S.; Frangione, V.; Ciani, O.; Tarricone, R.; Pappagallo, G. L. . A randomized, open-label, multicenter study of the efficacy and safety of intravesical hyaluronic acid and chondroitin sulfate versus dimethyl sulfoxide in women with bladder pain syndrome/interstitial cystitis. *Neurourology & Urodynamics* 2017;36:1178-86.
31. Gulpinar OE, B.; Kayis, A.; Gokce, M. I.; Suer, E. . Clinical comparison of intravesical hyaluronic acid and chondroitin sulfate therapies in the treatment of bladder pain syndrome/interstitial cystitis. *Neurourology & Urodynamics* 2018;37:257-62.
32. Hoaglin DC. Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. *Stat Med* 2016;35:485-95.
33. Hwang P, Auclair B, Beechinor D, Diment M, Einarson TR. Efficacy of pentosan polysulfate in the treatment of interstitial cystitis: a meta-analysis. *Urology* 1997;50:39-43.
34. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of clinical epidemiology* 1997;50:683-91.
35. Consilient Health. Pentosan polysulfate sodium for treating bladder pain syndrome [ID1364]. Clarification response; 2019.
36. Riedl C, Engelhardt P, Schwarz B. Treatment Costs of Bladder Pain Syndrome/Interstitial Cystitis in Austria: A Pharmacoeconomic Approach Following Current Guidelines. *Clinical Drug Investigation* 2013;33:737-42.
37. Hakimi Z, Houbiers J, Pedersini R, Vietri J. The Burden of Bladder Pain in Five European Countries: A Cross-sectional Study. *Urology* 2017;99:84-91.
38. NICE. Guide to the Methods of Technology Appraisal 2013. In; 2013.
39. Meunier A, Flint I, Sing J, Longworth L. Health-related utility and resource use of patients with interstitial cystitis/bladder pain syndrome (IC/BPS). London: PHMR; 2019.
40. van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, *et al.* Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value in Health* 2012;15:708-15.

41. The National Institute for Health and Care Excellence. Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018). London, UK: The National Institute for Health and Care Excellence,; 2018.
42. National life tables, UK - Office for National Statistics. In.
43. Improvement NHS. NHS reference costs 2017/18. National schedule of reference costs. In; 2018.
44. Curtis L, Burns A. Unit Costs of Health and Social Care 2017. In; 2017.
45. Department of Health. NHS Electronic Drug Tariff; 2018.
46. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. *Value Health* 2011;14:539-45.
47. Mims. Database of prescription and generic drugs, clinical guidelines | MIMS online. In.
48. StataCorp. Stata Statistical Software: Release 15. In: College Station, TX: StataCorp LLC; 2017.
49. Davis S, Stevenson M, Tappenden P, Wailoo A. NICE DSU Technical Support Document 15: Cost-effectiveness modelling using patient-level simulation. Sheffield: The University of Sheffield. ; 2014.
50. Economics CfBSiH. Excel macros for sampling from multivariate Normal and Dirichlet distributions. https://www.sheffield.ac.uk/chebs/software (Accessed

9 APPENDICES

Appendix 1: Unit costs and resource use for disease-related costs

	CS[1] unit cost	Reference in CS	Exact codes that were used	ERG preferred unit cost	Reference	Resource use	Average costs using ERG preferred costs	CS average costs
Community healthcare								
GP visit	£37.00	PSSRU (2017)[44]				[redacted]	[redacted]	[redacted]
Nurse visit	£10.50	PSSRU (2017)[44]	Uncertain of how per visit cost obtained- PSSRU 2017 reports cost per hour £36			[redacted]	[redacted]	[redacted]
Outpatients								
Outpatient visits	£125.00	NHS Reference Costs (2017/18)[43] Average of total outpatient attendances		£110.00	Outpatient visits to urology- service code 101 NHS Reference Costs (2017/18)[43]	[redacted]	[redacted]	[redacted]
Hospital admissions								

ITU LOS	£1,466.60	NHS Reference Costs (2017/18)[43] XC01Z Non-specific, General Adult Critical Care Patients Predominate. Unit cost is a weighted average of adult critical care, with 0-6+ organs supported.	NHS Reference Costs (2017/18)[43] XC01Z-XC07Z	£1,287.24	NHS Reference Costs (2017/18)[43] XC05Z- XC07Z Non-specific, General Adult Critical Care Patients Predominate. Unit cost is a weighted average of adult critical care, with **0-2 organs supported**.	[redacted]	[redacted]	[redacted]
General ward LOS	£327.00	NHS Reference Costs (2017/18)[43] Index- regular day or night admissions				[redacted]	[redacted]	[redacted]
Specialist ward LOS	£957.08	NHS Reference Costs (2017/18)[43] LB19E, LB19F & LB19G Ureteric or Bladder Disorders, without	Total HRGS- total LB19E, LB19F & LB19G			[redacted]	[redacted]	[redacted]

		Interventions						
Day case	£742.00	NHS Reference Costs (2017/18)[43] Index- (DC) day case	HRG of all types of day cases	£309.00	Day case- LB15E- Minor Bladder Procedures, 19 years and over NHS Reference Costs (2017/18)[43]	[redacted]	[redacted]	[redacted]
Gynaecology	£921.00	NHS Reference Costs (2017/18)[43] LB15E Service description: Gynaecology Currency description: Minor Bladder Procedures, 19 years and over	Elective inpatient			[redacted]	[redacted]	[redacted]
Accident and emergency								

Emergency	£244.93	NHS Reference Costs (2017/18)[43] LB15e, LB18Z Weighted average of all A&E visits directly linked to the bladder	Outpatient procedures-Accident and Emergency- LB15E and LB18Z weighted average			[redacted]	[redacted]	[redacted]
Ambulance	£252.00	NHS Reference Costs (2017/18)[43] ASS02 Currency description: See and Treat and Convey				[redacted]	[redacted]	[redacted]
Total							[redacted]	[redacted]
Difference							[redacted]	

Appendix 2: Corrections of unequivocal errors which were necessary to generate results for the scenario analyses for the company's revised base-case

The ERG identified an error in the company's original model which the company corrected in the VBA code (subroutine "DES_IC_BPS") used to run the base-case analysis in the company's revised model, submitted as part of the clarification responses[1] and received by the ERG on 13th February 2019. However, equivalent corrections were not carried through to the VBA code used to run their scenario analyses. The following corrections were necessary in order for the ERG to generate the scenario analyses provided in Table 26 and Table 27.

Correction 1: The ERG implemented the following change to correct the use of time to discontinuation without adjusting from months to years in the subroutines "SA_BI_admin", "SA_BSC_5y" ,"SA_BSC_6m", "SA_nonresp" and "SA_surery".

Original code;

```
CostsAccrued = CostsAccrued + (OtherCosts_resp * (T2TD - Range("time_response_check")))
DCostsAccrued = DCostsAccrued + (OtherCosts_resp * (Exp(T2TD * (0 - DRCosti)) - Exp(Range("time_response_check") * (0 - DRCosti))) / (0 - DRCosti))"
```

Corrected code;

```
CostsAccrued = CostsAccrued + (OtherCosts_resp * (T2TD - Range("time_response_check") / 12))
DCostsAccrued = DCostsAccrued + (OtherCosts_resp * (Exp(T2TD * (0 - DRCosti)) - Exp((Range("time_response_check") / 12) * (0 - DRCosti))) / (0 - DRCosti))
```

Correction 2: Typo in VBA code for running surgery scenario analysis where "Rsnd" was used instead of "Rand" in subroutine "SA_DES_surgery()"

Original code;

```
T2TD = Range("time_response_check") / 12 + (Exp(Constant_logn + Application.WorksheetFunction.Norm_S_Inv(Rsnd(n, 2)) * Shape_logn)) / 12
```

Corrected code;

```
T2TD = Range("time_response_check") / 12 + (Exp(Constant_logn + Application.WorksheetFunction.Norm_S_Inv(Rand(n, 2)) * Shape_logn)) / 12
```

Appendix 3: Technical appendix detailing methods for implementing ERG's exploratory analyses and additional sensitivity analyses

This appendix details the changes made by the ERG to implement exploratory analysis to create the ERG's preferred base case. All additional sensitivity analyses were conducted with all changes to exploratory analyses 1-7 implemented. All ERG model amendments were made to the company's revised model, submitted as part of the clarification responses[1] and received by the ERG on 13th February 2019.

Exploratory analysis 1: Use of ERG preferred survival analysis

A new worksheet has been added to the ERG revised model which contains STATA outputs using the ERGs preferred survival analysis data. In order to reproduce the STATA outputs copy columns A-C in worksheet "ERG survival curve" into a new STATA data editor. Run the STATA do file attached in appendix 4

Model parameters from the STATA output were pasted into cells 'K5:M13' in new worksheet "ERG survival curve".

Cells 'K5:M14' in new worksheet "ERG survival curve", were copy and pasted into worksheet "Model inputs", cells 'H119:J128'.

In worksheet "Model inputs", the following changes were made:

Set cell 'F120' equal to cell 'J120'. Named cell "Constant_exp_ERG".

Set cell 'F125' equal to cell 'J121'. Named cell "Constant_wei_ERG"

Set cell 'F126' equal to "=EXP(J122)". Named cell "Shape_wei_ERG"

Set cell 'F131' equal to cell 'J127'. Named cell "Constant_logn_ERG"

Set cell 'F132' equal to "=EXP(J128)". Named cell "Shape_logn_ERG"

In VBA module "DES", a new variable was defined to switch to ERG preferred inputs for time to discontinuation using the following code:

"*Dim T2TD_ERG_flag As Integer*"

"*T2TD_ERG_flag = Range("T2TD_ERG_flag").Value*"

Added "*If T2TD_ERG_flag = 0 Then*" before the following VBA code:

```
Constant_exp = Range("Constant_exp")
 Constant_wei = Range("Constant_wei")
 Shape_wei = Range("Shape_wei")
 Constant_logn = Range("Constant_logn")
 Shape_logn = Range("Shape_logn")
```

Added the following code immediately after the above code:

ElseIf T2TD_ERG_flag = 1 Then
Constant_exp = Range("Constant_exp_ERG")
Constant_wei = Range("Constant_wei_ERG")
Shape_wei = Range("Shape_wei_ERG")
Constant_logn = Range("Constant_logn_ERG")
Shape_logn = Range("Shape_logn_ERG")
Else
MsgBox "survival parameters not selected"
End If

These amendments can be implemented by entering a value of "1" into worksheet "ERG options" cell B14 and selecting 'Click to apply default values to model' in worksheet "Control", cell C2 prior to running the model.

Exploratory analysis 2: Switch to 6-weekly dosing for first line BIs after first year of treatment and 6-weekly for all 2nd line BIs

In VBA module "DES", a new variable was defined used to switch on increased spacing of bladder instillation doses post year 1 using the following code:

"*Dim Dosing_flag As Integer*"

Dosing_flag = Range("dose_spacing_flag").Value

In 'Case 4: Next event is drug administration' section of DES module, under:

'If DrugAdministrationCount < 4 Then
T2DrugAdministration = T2DrugAdministration + (1 / 52)'

The following new code was added to allow for increased dosing after 1 year:

"ElseIf DrugAdministrationCount >= 15 And Dosing_flag = 1 Then T2DrugAdministration = T2DrugAdministration + (6 / 52)"

The value in worksheet "Model inputs", cell D56, the formula has been revised to the following to allow for 6 weekly dosing for all second line BI treatments:

"=IF(dose_spacing_flag=0,(D46+D51)*13,(D46+D51)*(52/6))"

These amendments can be implemented by entering a value of "1" into worksheet "ERG options" cell B15 and selecting 'Click to apply default values to model' in worksheet "Control", cell C2 prior to running the model.

Exploratory analysis 3: Use regression for utility based on ICSI scores which excludes term for prior usage of BI

Data from the regression analysis excluding recent/current usage of bladder instillations has been copied from the company's additional document 'ID1364_PPS_utilities_generation_report_AIC', provided during clarification process,[1] into worksheet 'Response & Utility data' in cells E61:F95
Within the worksheet 'Response & Utility data', the following cells were amended:

- The formula in cell H50 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G50+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G50+$F$66))))"

- The formula in cell H51 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G51+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G51+$F$66))))"

- The formula in cell H52 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G52+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G52+$F$66))))"

- The formula in cell H53 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G53+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G53+$F$66))))"

- The formula in cell H54 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G54+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G54+$F$66))))"

- The formula in cell H55 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G55+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G55+$F$66))))"

- The formula in cell H56 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G56+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G56+$F$66))))"

- The formula in cell H57 was amended to "=IF('ERG options'!$B$16=0,1/(1+EXP(-1*($C$78+$C$62*G57+$C$68))),1/(1+EXP(-1*($F$75+$F$62*G57+$F$66))))"

- The formula in cell I50 was amended to "=IF('ERG options'!$B$16=0,$C$97+$C$81*G50+$C$87,$F$94+$F$78*G50+$F$83)"

- The formula in cell I51 was amended to "=IF('ERG options'!$B$16=0,$C$97+$C$81*G51+$C$87+$C$93,$F$94+$F$78*G51+$F$83)"

- The formula in cell I52 was amended to “IF('ERG options'!$B$16=0,$C$97+$C$81*G52+$C$87,$F$94+$F$78*G52+$F$83)”

- The formula in cell I53 was amended to “=IF('ERG options'!$B$16=0,$C$97+$C$81*G53+$C$87,$F$94+$F$78*G53+$F$83)”

- The formula in cell I54 was amended to “=IF('ERG options'!$B$16=0,$C$97+$C$81*G54+$C$87+$C$93,$F$94+$F$78*G54+$F$83)”

- The formula in cell I55 was amended to “=IF('ERG options'!$B$16=0,$C$97+$C$81*G55+$C$87+$C$93,$F$94+$F$78*G55+$F$83)”

- The formula in cell I56 was amended to “=IF('ERG options'!$B$16=0,$C$97+$C$81*G56+$C$87,$F$94+$F$78*G56+$F$83)”

- The formula in cell I57 was amended to “=IF('ERG options'!$B$16=0,$C$97+$C$81*G57+$C$87,$F$94+$F$78*G57+$F$83)”

These amendments can be implemented by entering a value of “1” into worksheet “ERG options” cell B16 and selecting ‘Click to apply default values to model’ in worksheet “Control”, cell C2

Exploratory analysis 4: Use of a lifetime horizon

Value in worksheet “Model inputs”, cell D13 was replaced with “*=IF('ERG options'!B17=0,20,100)*”
This amendment can be implemented by entering a value of “1” into worksheet “ERG options” cell B17 and selecting ‘Click to apply default values to model’ in worksheet “Control”, cell C2 prior to running the model.

Exploratory analysis 5: Return to baseline utilities and costs for non-responders when BSC is second line option (PPS versus BSC scenario only)

- In VBA module “DES”, a new variable was defined to switch on the option for patients receiving BSC to return to baseline costs and utilities, by adding the following code:

“*Dim return2baseline_flag As Integer*”
“return2baseline_flag = Range("return2baseline").Value
If return2baseline_flag = 1 Then Sheet5.Range("Selected_2nd") = 2”

- To set costs to return baseline values for non-responder, the below code:

"bladcost_annual = Choose(Range("Selected_2nd"), Range("blad_cost_annual") + Range("blad_cost_2nd"), Range("bsc_cost_nonresp_no2nd"))"

Is replaced with the following:

"Dim temp As Double
If return2baseline_flag = 0 Then
temp = Range("bsc_cost_nonresp_no2nd")
Else
temp = Range("bsc_cost_pre")
End If
bladcost_annual = Choose(Range("Selected_2nd"), Range("blad_cost_annual") + Range("blad_cost_2nd"), temp)"

- To set utilities to baseline values for non-responders, the following code under 'If i=1 Then':
  "Utility_nonresp = Choose(Range("Selected_2nd"), Range("blad_utility_2nd"), Range("elmiron_utility_nonresp_no2nd"))"

  Is replaced with the following:
  " If return2baseline_flag = 0 Then
  temp = Range("elmiron_utility_nonresp_no2nd")
  Else
  temp = Range("elmiron_utility_pre")
  End If
  Utility_nonresp = Choose(Range("Selected_2nd"), Range("blad_utility_2nd"), temp)"

- To set utilities to baseline values for non-responders, the following code under 'If i=2 Then':
  "Utility_nonresp = Choose(Range("Selected_2nd"), Range("blad_utility_2nd"), Range("bsc_utility_nonresp_no2nd"))"

  Is replaced with the following:
  "*If return2baseline_flag = 0 Then*
  temp = Range("bsc_utility_nonresp_no2nd")
  Else
  temp = Range("bsc_utility_pre")
  End If
  Utility_nonresp = Choose(Range("Selected_2nd"), Range("blad_utility_2nd"), temp)"

- To set utilities to baseline values for non-responders, the following code under 'If i=3 Then': "Utility_nonresp = Choose(Range("Selected_2nd"), Range("blad_utility_2nd"), Range("blad_utility_nonresp"))"

  Is replaced with the following:
  "If return2baseline_flag = 0 Then
  temp = Range("blad_utility_nonresp")
  Else
  temp = Range("blad_utility_pre")
  End If
  Utility_nonresp = Choose(Range("Selected_2nd"), Range("blad_utility_2nd"), temp)"

These amendments can be implemented by entering a value of "1" into worksheet "ERG options" cell B18 and selecting 'Click to apply default values to model' in worksheet "Control", cell C2

Exploratory analysis 6: Switch off receding baseline response for BSC (PPS versus BSC scenario only as already implemented in PPS versus BI base case)
Set 'Placebo effect receding' switch on worksheet 'Model inputs' to "NO".

Exploratory analysis 7: Use of log-normal function to model time to discontinuation
Apply all changes from ERG exploratory analysis 1.
Set drop down selection on worksheet 'Model inputs', cell C110 to "Lognormal"

Additional sensitivity analysis 1: All costs based on baseline ISCI scores removing the relationship between response to treatment and costs
Within worksheet "Cost & Survival data", the following cells were amended:

- The formula in Cell D38 was amended to "=EXP($C$58+$C$50*IF('ERG options'!B26=0,'Response & Utility data'!D36,'Response & Utility data'!D36)+$C$54)"

- The formula in Cell D39 was amended to "=EXP($C$58+$C$50*IF('ERG options'!B26=0,'Response & Utility data'!G44,'Response & Utility data'!D36)+$C$54)"

- The formula in Cell D40 was amended to "=EXP($C$58+$C$50*IF('ERG options'!B26=0,'Response & Utility data'!H44,'Response & Utility data'!D36)+$C$54)"

- The formula in Cell D42 was amended to "=EXP($C$58+$C$50*IF('ERG options'!B26=0,'Response & Utility data'!G45,'Response & Utility data'!D36)+$C$54)"

- The formula in Cell D43 was amended to "=EXP($C$58+$C$50*IF('ERG options'!B26=0,'Response & Utility data'!H45,'Response & Utility data'!D36)+$C$54)"

- The formula in Cell D45 was amended to "=EXP($C$58+$C$50*IF('ERG options'!B26=0,'Response & Utility data'!G46,'Response & Utility data'!D36)+$C$54)"

- The formula in Cell D45 was amended to "=EXP($C$58+$C$50*IF('ERG options'!B26=0,'Response & Utility data'!H46,'Response & Utility data'!D36)+$C$54)"

These amendments can be implemented by entering a value of "1" into worksheet "ERG options" cell B26 and selecting 'Click to apply default values to model' in worksheet "Control", cell C2

Additional sensitivity analysis 2: Explore different baseline response rates through changes to response rate of BSC based on upper and lower confidence intervals reported in the literature.
A value of 5% was added to worksheet "Response & Utility data", cell K27.
A value of 32% was added to worksheet "Response & Utility data", cell L27.

Formula in worksheet "Response & Utility data", cell J24, was replaced with the following formula: "=IF('ERG options'!B27=0,G27,IF('ERG options'!B27=-1,'Response & Utility data'!K27,'Response & Utility data'!L27))"

These amendments can be implemented by entering a value of "-1" for lower response rate of 5% or a value of "1" for upper response rate of 32% into worksheet "ERG options" cell B27 and selecting 'Click to apply default values to model' in worksheet "Control", cell C2.

Additional sensitivity analysis 4 & 5: Urology outpatient cost used for administration of BIs & 80% self-administer BIs (PPS versus BI only)
A value of £151.05 was added to worksheet "Cost & Survival data", cell C16.

Value in worksheet "Control", cell C15, was replaced with the following formula:
*"=IF((AND('ERG options'!B25=0,'ERG options'!B28=0)),'Cost & Survival data'!C15,(IF('ERG options'!B25=1,'Cost & Survival data'!C16,('Cost & Survival data'!C15-('Cost & Survival data'!C15*0.8)))))*"

Amendments to sensitivity analysis 4 can be implemented by entering a value of “1” into worksheet “ERG options” cell B28.

Amendments to sensitivity analysis 5 can be implemented by entering a value of “1” into worksheet “ERG options” cell B29.

Appendix 4: STATA code used to run ERG's survival analysis of time to treatment discontinuation

```
stset time,failure(failure) id(id)

streg, dist(exponential)
estat ic
predict mean_time_ex, mean time
predict median_time_ex, time

streg, dist(weibull)
estat ic
predict mean_time_we, mean time
predict median_time_we, time

streg, dist(lognormal)
estat ic
predict mean_time_ln, mean time
predict median_time_ln, time

streg, dist(gompertz)
estat ic
predict median_time_gpz, time

streg, dist(loglogistic)
estat ic
predict mean_time_lgl, mean time
predict median_time_lgl, time

sts graph,xlabel(0(6)60) risktable
```