


BMJ Open Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis

Andrew Bryant ¹, Michael Grayling,¹ Shaun Hiu,¹ Ketankumar Gajjar,² Eugenie Johnson,¹ Ahmed Elattar,³ Luke Vale,¹ Dawn Craig,¹ Raj Naik⁴

To cite: Bryant A, Grayling M, Hiu S, *et al*. Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis. *BMJ Open* 2022;**12**:e060183. doi:10.1136/bmjopen-2021-060183

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-060183>).

Received 14 December 2021
Accepted 03 August 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Andrew Bryant;
andy.bryant@ncl.ac.uk

ABSTRACT

Objectives We consider expert opinion and its incorporation into a planned meta-analysis as a way of adjusting for anticipated publication bias. We conduct an elicitation exercise among eligible British Gynaecological Cancer Society (BGCS) members with expertise in gynaecology.

Design Expert elicitation exercise.

Setting BGCS.

Participants Members of the BGCS with expertise in gynaecology.

Methods Experts were presented with details of a planned prospective systematic review and meta-analysis, assessing overall survival for the extent of excision of residual disease (RD) after primary surgery for advanced epithelial ovarian cancer. Participants were asked views on the likelihood of different studies (varied in the size of the study population and the RD thresholds being compared) not being published. Descriptive statistics were produced and opinions on total number of missing studies by sample size and magnitude of effect size estimated.

Results Eighteen expert respondents were included. Responders perceived publication bias to be a possibility for comparisons of RD <1 cm versus RD=0 cm, but more so for comparisons involving higher volume suboptimal RD thresholds. However, experts' perceived publication bias in comparisons of RD=0 cm versus suboptimal RD thresholds did not translate into many elicited missing studies in Part B of the elicitation exercise. The median number of missing studies estimated by responders for the main comparison of RD<1 cm versus RD=0 cm was 10 (IQR: 5–20), with the number of missing studies influenced by whether the effect size was equivocal. The median number of missing studies estimated for suboptimal RD versus RD=0 cm was lower.

Conclusions The results may raise awareness that a degree of scepticism is needed when reviewing studies comparing RD <1 cm versus RD=0 cm. There is also a belief among respondents that comparisons involving RD=0 cm and suboptimal thresholds (>1 cm) are likely to be impacted by publication bias, but this is unlikely to attenuate effect estimates in meta-analyses.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ In our elicitation exercise, designed in collaboration with senior gynae-oncologists, the number of respondents (n=18) was sufficient to provide a solid basis for meaningful conclusions to be drawn in an area of uncertainty.
- ⇒ Part A of the elicitation identifies areas where publication bias is of concern, but the questions asked do not provide an indication of the direction of any bias.
- ⇒ Therefore, in Part B of our elicitation exercise, we collected information that would enable any planned meta-analysis estimates to be adjusted for the anticipated impact of publication bias.
- ⇒ The approach adopted is inexpensive and easy to design and administer and did not rely on any contact with participants, who were able to complete at their own convenience.
- ⇒ However, answers given by the experts to open-ended questions were prone to an 'extreme answer bias'.

INTRODUCTION

Residual disease (RD) after upfront primary debulking surgery (PDS) for advanced epithelial ovarian cancer (EOC) is believed to be a key determinant of overall survival (OS). A recent prognostic factor systematic review protocol aims to demonstrate the superiority in terms of OS of the complete removal of RD in advanced EOC compared with leaving macroscopic disease (that is, the surgeon leaving some visible disease).¹

However, much of the evidence in this area comes from small and/or retrospective studies. Relying on such studies to draw conclusions may be unsound. One reason for this relates to possible publication biases, which may be more pronounced for small,

retrospective evaluations. Publication bias can arise when the publication of research findings depends on the nature and direction of the results. It is more likely in smaller and retrospective studies than for larger randomised controlled trials.²⁻⁶ Small studies might be underpowered and, furthermore, null findings might be due to deficiencies in the study design and conduct. Hence, including these studies might not lead to an appropriate adjustment of meta-analysis estimates. This is why we planned to include studies with a minimum sample size of 100 patients in the systematic review.

Therefore, given the nature of the evidence base, publication bias could be hypothesised to lead to a bias in favour of more complete removal of RD as described below.

Small and retrospective studies are also prone to other biases, particularly selection bias, (ie, systematic differences between groups in terms of baseline characteristics) compared with randomised trials.^{7 8} Furthermore, all study designs may suffer from inadequacies of study conduct, such as deficiencies in blinding, high attrition and so on.^{9 10} Again, these problems are potentially exacerbated for smaller retrospective studies.

As alluded to above, publication and other reporting biases^{11 12} can have serious consequences to research and impact on summary of findings and recommendations in guidelines.^{13 14} If it is suspected that publication bias is highly plausible, this may make the effect estimates obtained from meta-analyses uncertain and potentially unreliable. This is a concern when considering the results of the systematic review assessing OS for RD after PDS in advanced EOC.¹

In this review, the data underpinning the estimates will be derived from the further analysis of data collected to address other research questions. Post hoc analyses of data collected to address other questions and secondary analyses of past medical records do not have to be prespecified anywhere, so there is a strong threat of data dredging. Therefore, the reporting of such data for individual studies may depend on the significance of their findings. For example, it is possible that only analyses producing 'significant' findings will be published. Thus, any meta-analysis may overestimate the effect of complete cytoreduction. This may be true even if many of the non-reported studies are small, as their cumulative impact on the meta-analysis may have a substantial overall effect.

Exploration of publication bias is an important part of a robust systematic review and should always be considered. At present, there is no consensus on a standard approach for identifying and adjusting for publication bias, although some methods, particularly around identification, do exist. Reduction of publication bias can be achieved by adherence to good review practice, such as a thorough search of grey literature.¹⁵⁻¹⁷ Post hoc statistical approaches such as funnel plots,¹⁸ trim and fill,^{19 20} and file drawer number²¹ could also be used. Furthermore, when there is evidence for publication bias or this bias is highly suspected, selection models^{22 23} might be used

to investigate how the results of a meta-analysis may be affected by publication bias. However, these usually require a large number of included studies in the analysis^{12 24} and any adjustment generally requires an assumption of the underlying selection model.^{12 22}

A potentially more practical approach is to incorporate external information into the meta-analysis. This external information could be gathered from various sources and incorporated using a Bayesian framework.²⁵⁻²⁷ However, this approach would only be useful if the external information is obtained from a reliable source. This final point is the focus of our study, as we propose an approach that has hitherto received little attention in meta-analyses: the consideration of expert opinion and the incorporation of their views and opinions into the meta-analysis to inform the adjustment. We do this by conducting an elicitation exercise among eligible British Gynaecological Cancer Society (BGCS) members (based on a pertinent job title and expertise in gynaecology) to identify their expert opinions on the potential nature and extent of publication bias in a planned prospective systematic review and meta-analysis assessing OS in RD after PDS for advanced EOC.¹ The elicitation exercise relates to the conduct of the planned systematic review, where the findings from this exercise will be used to adjust the proposed meta-analyses for any perceived publication bias.

In the elicitation exercise, we ask participants to account for: (1) the sort of studies that have been conducted but not published; (2) the plausible magnitude and direction of any publication bias; and (3) possible explanations for why and how the publication bias occurs. These data could be used to adjust the results for publication bias in our planned meta-analysis assessing OS in RD thresholds after primary surgery for EOC.

METHODS

Case study

This research involved human participants outside of a study or trial setting. The elicitation exercise did not require ethical approval because it was sent to BGCS members and participation was optional. Information about any expert that participated in the elicitation exercise was kept confidential.

Participants were given details of a planned prospective systematic review and meta-analysis assessing OS for the extent of excision of RD (see online supplemental appendix 1). This will include data from studies or case series of 100 or more patients that include a concurrent comparison of different RD thresholds after primary surgical intervention in adult women with advanced EOC. The outcome of interest was OS for different categories of RD.

For the purposes of the case study, participants were told that bibliographic databases up to January 2020 were searched for pertinent data, so that they had a cut-off for their responses to each scenario. Participants were made aware that two review authors would independently

abstract data and assess risk of bias and, where possible, that the data would be synthesised in a meta-analysis. Full details of the methodology used in the review is provided in a Cochrane systematic review¹ and a summary of inclusion criteria is given in the elicitation exercise in online supplemental appendix 1.

The review objective is to assess the impact of RD after upfront and interval debulking surgery on survival outcomes. However, the focus of this paper and the elicitation exercise was OS in different RD thresholds after upfront primary surgery.

Design of elicitation exercise

The purpose of the elicitation exercise was to ask respondents for their opinions on the likelihood of studies not being published. Thereafter, we asked for their opinions on several different scenarios, all of which related to the likelihood of different studies not being published. These unpublished studies varied by both size of the study population and the impact of the RD threshold as a prognostic factor for OS.

The elicitation exercise was designed in consultation with four gynaecologists, to help ensure a sufficiently detailed level of explanation was provided regarding the purpose of the exercise, along with clear descriptions of the methodology and rationale. Visual examples were used to make what was being asked of respondents as transparent as possible.

Usually, expert opinions are elicited either directly using interview methods or via an elicitation exercise. In either case, opinions potentially need only be provided by as few as four experts.^{28 29} However, it is advised to use more experts to give the results more generalisability and allow for the potential of a broader range of views.^{30 31} Any widespread disagreement among experts can be reflected in the uncertainty of elicited estimates; all that is fundamental is that respondents have extensive knowledge and expertise in the area of interest.

The elicitation exercise consisted of three parts: A, B and C (online supplemental appendix 1 provides an example of the elicitation exercise). Part A adopted an existing method of elicitation,³⁰ while Part B used a de novo tool designed to provide a way of obtaining an estimate of the number of missing studies from a meta-analysis. Respondents also indicated the size of these missing studies, which can be used to calculate the magnitude of effect in the form of a HR with 95% CI. Parts A and B are described in more detail below. Part C was used to gauge the attitudes of the respondent cohort about reporting biases more generally and is not reported here.

To assist respondents in answering questions in Parts A and B, we provided brief guidance on the interpretation of commonly reported statistics from survival models (see introductory section of 'Expert elicitation' in online supplemental appendix 1).

Expert elicitation Part A

This part comprised one question (Q1) and attempted to assess publication bias by asking respondents about their

views on the chance of publication for comparisons of different macroscopic RD thresholds (RD>0cm) versus the reference comparator of complete cytoreduction (removal of tumour so that there was no visible disease with the naked eye, RD =0cm). Specifically, for each comparison the sample size of the hypothetical study was varied between a minimum sample size (n=100) (which was part of the inclusion criteria in the planned review) and a maximum sample size (this maximum was based on observed sizes in the meta-analysis of included studies in an initial scope of the results up to January 2020).

Responders were then asked to assign a probability that a study reporting a given comparison with a given sample size would be published on a scale of 0 (no chance of publication) to 100 (certainly published). Other characteristics of the hypothetical study followed the inclusion criteria set out in the systematic review protocol by Bryant *et al.*¹ These have been summarised above and are reported in online supplemental appendix 1.

Expert elicitation Part B

Part B consisted of three broad questions and aimed to obtain the opinion of respondents on the estimated number of conducted-but-unpublished studies that might exist. For each question, participants were asked to consider a particular macroscopic RD threshold and compare it with RD =0cm: Q2 (RD <1 cm vs RD =0cm); Q3 (RD >1 cm vs RD =0cm); and Q4 (RD >2cm vs RD =0cm). Subsequently, participants were asked on a Likert Scale from 1 (not likely at all) to 5 (extremely likely), the likelihood that relevant studies that either favoured macroscopic disease, or studies that found no statistically significant difference ($p>0.05$) in survival between macroscopic disease and RD =0cm, would not be published.

Next, respondents were asked to give an estimate of how many studies of a certain size and magnitude of effect might be unpublished, along with a rationale for their answer. The sample size of unpublished studies was varied in increments of 100 from 100 to >500. The effect size, reported as the adjusted HR, was likewise varied in decrements of 0.1, between 1 and ≤ 0.5 . In total, respondents were asked to think about the number of unpublished studies for 36 different hypothetical combinations of sample size and effect size. The questions were repeated for scenarios involving suboptimal RD thresholds (>1 cm and >2 cm) compared with RD =0 cm (See Q3-4 of elicitation exercise in online supplemental appendix 1).

The responses to the questions in Part B could be used to adjust the overall effect estimate from observed studies when data from unobserved studies are added.

Data collection and sampling

The elicitation exercise was vetted by the BGCS Survey panel; their helpful suggestions were incorporated and a link to the finalised elicitation exercise using Qualtrics was distributed to members via email by the BGCS administrator. BGCS have established guidelines for circulation of online surveys via the membership email

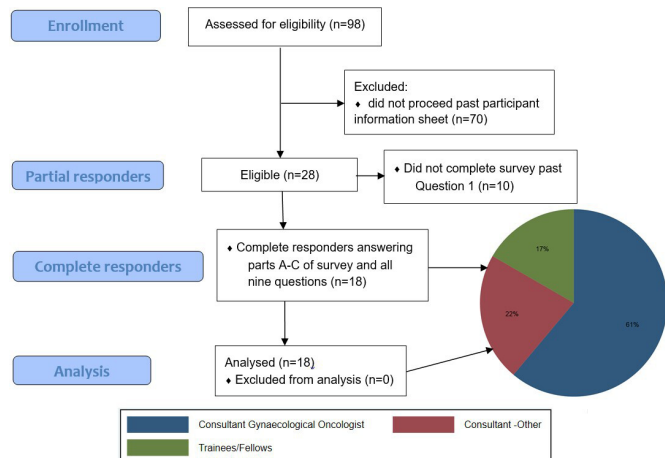


Figure 1 Elicitation exercise flow diagram.

directory, which were followed in our elicitation exercise and are available on request to the BGCS. The link to the elicitation exercise was open from 13 August 2020 to 26 October 2020 and two reminders were sent out. Study participation was voluntary and potential respondents were informed that the results of the elicitation would inform a publication. All acknowledgements are given with the consent of responders; all open-text responses provided have been anonymised and we have explicitly excluded cross-tabulation by job title, as this may have compromised the anonymisation.

Data analysis

The responses of the elicitation exercise are summarised using descriptive statistics. Further details are reported in online supplemental appendix 1. For the responses to Part B, we also provide in online supplemental appendix 2 an example of how the responses could be used to form an overall estimate of the total number of missing studies by sample size and magnitude of effect size for each question, reported as a HR and 95% CI. All analyses were conducted in StataIC V.15.³²

Patient and public involvement

None.

RESULTS

Characteristics of respondents

The elicitation exercise was sent to all 455 BGCS members at the time, with over 80% being eligible to complete. A total of 98 BGCS members opened the link for the exercise and 28 proceeded past the participant information sheet. Of these, 18 respondents fully completed the elicitation exercise, and their responses are reported below. The remaining 10 participants did not adequately contribute to the exercise to be included in analysis (figure 1).

The distribution of expertise of completers of the exercise is also presented in figure 1. Most responders were consultant gynaecological oncologists (11/18; 61%) or subspecialist consultants (4/18; 22%). The median time to complete the exercise was 18 min (IQR 16–27 min) with a range of 8–61 min. The mean completion time was 23 min (SD 14 min).

Part A: Probability estimates that a study with minimum and maximum specified sample sizes is published for different macroscopic RD disease versus RD =0 cm

Table 1 shows the perceived probability that a study is published based on its sample size for the comparison of different RD thresholds (all compared with RD =0 cm). Responses suggest that publication bias may be quite likely in studies where the sample size was just 100. For example, responders suggest they thought there was less than a 60% chance that a comparison of RD <1 cm versus RD =0 cm would be reported for a study with a sample size of 100 participants.

Overall, there was widespread variation in the results, indicating that some responders thought the probability of publication was much higher than others (range 0%–100%). Responders appeared to indicate that the probability of publication was lowest for comparisons involving greater macroscopic disease volume (largest elicited median probability 20% (IQR 10%–75%) in macroscopic disease involving

Table 1 Summary statistics of responders' perceived chance (probability) of publication for studies of given sample size for residual disease thresholds compared with microscopic disease (0 cm)

Versus 0 cm RD threshold	% for n minimum (n=100)			% for n maximum		
	Mean (SD)	Median (IQR)	Observed range	Mean (SD)	Median (IQR)	Observed range
< 1 cm	57 (31.2)	55 (30–80)	0–100	95 (6.1)	99.5 (90–100)	80–100
> 0 cm	49 (33.6)	50 (20–80)	0–95	77 (25.5)	80 (70–99)	0–100
1–2 cm	48 (32.1)	50 (20–70)	0–100	58 (34.1)	72.5 (30–80)	0–100
< 2 cm	50 (36.6)	50 (10–85)	0–100	58 (36.6)	65 (20–90)	0–100
> 1 cm	49 (34.4)	45 (20–90)	0–100	85 (19.3)	95 (75–99)	40–100
> 2 cm	38 (36.6)	20 (10–75)	0–100	47 (37.9)	30 (15–80)	0–100
1–5 cm	29 (33.2)	10 (0–50)	0–95	42 (38.2)	27.5 (5–80)	0–100
> 5 cm	23 (34.4)	3.5 (0–50)	0–95	35 (40.7)	10 (0–80)	0–100

N maximum varies for different RD =0 cm versus RD threshold comparisons. For RD <1 cm and RD >1 cm, n=1000. For RD >0 cm, n=625, For RD 1–2 cm, n=210. The remainder are n=250.
RD, residual disease;

Table 2 Responders' perceived likelihood of publication bias in comparisons of near optimal (<1 cm) and suboptimal (>1/2 cm) versus complete cytoreduction (0 cm)

Perceived likelihood of publication bias	RD <1 cm vs 0 cm		RD >1 cm vs 0 cm		RD >2 cm vs 0 cm	
	N	%	N	%	N	%
Not likely at all (1)	1	5.5	10	55.5	15	83.5
Somewhat likely (2)	5	28	2	11	1	5.5
Quite likely (3)	8	44.5	3	17	0	0
Very likely (4)	2	11	2	11	1	5.5
Extremely likely (5)	2	11	1	5.5	1	5.5

RD, residual disease.

RD >2 cm vs RD =0 cm and as low as 3.5% (IQR: 0%–50%) for RD >5 cm vs RD =0 cm).

Respondents also indicated that there was potential for publication bias in some comparisons when studies had larger sample sizes. However, responders appeared to dismiss the threat of publication bias for comparisons of RD <1 cm versus RD =0 cm and RD >1 cm versus RD =0 cm. Mean and median probabilities were higher and close to 100%, indicating that respondents were highly certain that a study would be published. Comparisons involving higher volume suboptimal RD (greater macroscopic disease volume) versus RD =0 cm were considered to have a low probability of being published for larger studies (the largest elicited median probability was 30% (IQR: 15%–80%) in macroscopic disease involving RD >2 cm and the probability was much less for RD 1–5 cm and RD >5 cm). This was consistent with the results for smaller studies.

Part B: Perceived likelihood of publication bias and estimation of missing studies

Table 2 shows that most responders acknowledged that the likelihood of publication bias is 'somewhat' or 'quite' likely (72.5%) in the comparison of RD <1 cm with RD =0 cm, with only one responder (5.5%) thinking it was not likely at all. This view was completely reversed for comparisons involving suboptimal RD >1 cm with RD =0 cm, where most responders thought publication bias was 'not likely at all'.

The mean and median numbers of missing studies estimated by responders for comparison of RD <1 cm

versus RD =0 cm was 17 (SD 16.5) and 10 (IQR 5–20), respectively (table 3). The average number of estimated missing studies was lower for the comparisons involving suboptimal macroscopic disease volume (RD thresholds that are >1 cm). The mean and median numbers of missing studies estimated by responders for the comparison of RD >1 cm versus RD =0 cm was 8.6 (SD 12.9) and 5 (IQR 0–10), respectively (table 3). The mean number of missing studies estimated by responders for the comparison of RD >2 cm versus RD =0 cm was 6 (SD 13.2) and median was 0.5 (IQR 0–5) (table 3).

Table 4 and the tables in online supplemental appendix 3 and 4 show that, in the opinion of respondents, the number of studies that might be missing may be influenced by the effect size of those missing studies detected. For example, for the comparison of RD <1 cm versus RD =0 cm, on average 9.4 of the 17 studies would be associated with an HR of 1. As the HR increased, fewer studies were felt to be missing such that, when the detected HR was 0.5, the average number of studies felt to be missing was less than 1. Considering all the studies that were felt to be missing by respondents, a weighted average HR was estimated. This weighted average HR of the effect size from the missing studies was 0.83 (95% CI 0.77 to 0.90) for the comparison of RD <1 cm compared with RD =0 cm. This HR was calculated based on a total of 3906 participants in the estimated missing studies and 2500 deaths given a 5-year survival rate of 36% (table 4).

Table 3 Summary statistics of responders' perceived likelihood of publication bias in comparisons of near optimal (<1 cm) and suboptimal (>1/2 cm) versus complete cytoreduction (0 cm)

Summary statistics (Scale 1–5)	RD <1 cm vs 0 cm			RD >1 cm vs 0 cm			RD >2 cm vs 0 cm		
	Mean (SD)	Median (IQR)	Range	Mean (SD)	Median (IQR)	Range	Mean (SD)	Median (IQR)	Range
Overall score of perceived likelihood of publication bias (n=18)	2.94 (1.1)	3 (2–3)	1–5	2 (1.3)	1 (1–3)	1–5	1.4 (1.1)	1 (1–1)	1–5
Total estimated missing studies (n)	17.8 (16.5)	10 (5–20)	0–50	8.6 (12.9)	5 (0–10)	0–50	6.2 (13.2)	0.5 (0–5)	0–50

RD, residual disease.

Table 4 Breakdown of distribution of size and magnitude of elicited unpublished studies of near-optimal RD <1 cm versus complete cytoreduction (0 cm)

n=321 (n=17.8)		Estimated effect size					
Assumed 5-year survival: 36%		HR=1	HR=0.9	HR=0.8	HR=0.7	HR=0.6	HR≤0.5
Size of studies missed that could have been included in the analysis	Sample size	RD <1 cm and 0 cm are the same	10% less chance of mortality favouring RD <1 cm	20% less chance of mortality favouring RD <1 cm	30% less chance of mortality favouring RD <1 cm	40% less chance of mortality favouring RD <1 cm	≥50% less chance of mortality favouring RD <1 cm
	n<100	Study excluded					
	n=100	122.08*	19.12	22.7	1.34	2.14	1.14
	n=200	25.08	11.12	12.62	4.38	2.18	2.18
	n=300	6.04	4.04	1.04	2.04	0	0
	n=400	10.37	9.37	9.37	9.37	9.37	9.37
	n=500	1.04	1.04	3.04	1.04	0	0
	n>500	5.08	4.04	4.04	3.04	1.04	1.04
Total studies† (mean)		169.7 (9.4)	48.7 (2.7)	52.8 (2.9)	21.2 (1.2)	14.7 (0.8)	13.7 (0.8)
Effective n‡ (mean)		26 879 (1493.3)	12 141 (674.5)	12 899 (716.6)	7790 (432.8)	5048 (280.4)	4948 (274.9)
Effective d§ (mean)		17 203 (956)	7770 (432)	8255 (459)	4986 (277)	3231 (179)	3167 (176)
SElogHR (√(4/d))¶		0.065	0.096	0.093	0.120	0.149	0.151
95% CI for HR**		0.88–1.14	0.75–1.09	0.67–0.96	0.55–0.89	0.45–0.80	0.37–0.67
Elicited estimate††		HR 0.83 (95% CI 0.77 to 0.90), logHR –0.19 SElogHR 0.04 (n=3906, d=2500)					

*Number of studies given in the breakdown were rescaled in three respondents to correspond to the total number estimated. Therefore, any non-integer numbers in the table are due to this rescaling.

†Absolute number of estimated missing studies elicited from responders with mean (simply absolute number divided by 18 (number of responders)) given in parentheses.

‡Absolute number of estimated missing participants elicited based on total studies with mean given in parentheses.

§Absolute number of deaths estimated from number of participants assuming 5-year survival rate of 36% with mean in ().

¶Approximation of the SE of the log HR using formula derived by Parmar,⁴⁶ namely the square root of 4 divided by mean number of deaths.

**95% CI for HR calculated using $\log HR \pm 1.96$ multiplied by SE of log HR then transforming back by taking the exponential.

††Elicited HR with 95% CI using mean responses for all aggregated effect sizes.

RD, residual disease.

Similarly, the mean number of missing studies estimated by responders for comparison of RD >1 cm versus RD =0 cm was 8.6 (table 3). The weighted average HR of the missing studies estimated HR 0.77 (95% CI 0.70 to 0.85); this was estimated using the same approach as described above, as reported in online supplemental appendix 3. The mean number of missing studies estimated by responders for comparison of RD >2 cm versus RD =0 cm was 6.2 (table 3). The weighted average HR was estimated to be 0.79 (95% CI 0.71 to 0.89; see online supplemental appendix 4 for more details of the data).

A further analysis of results by the strength of responders' opinions as to the likelihood of publication bias was conducted. Calculating an overall HR and 95% CI for missing studies based on responders in these likelihood of publication subgroups ('not likely at all', 'somewhat likely', 'quite likely', 'very or extremely likely') led to an estimated HR of 0.90 (95% CI 0.79 to 1.03) for comparison of RD <1 cm versus RD =0 cm (table 5). These analyses were not repeated for comparisons of RD >1 cm versus RD =0 cm and RD >2 cm versus RD =0 cm,

as the opinions of responders shifted towards a general feeling that publication bias was 'not likely at all'. The range in the estimated number of conducted but unpublished studies according to RD <1 cm versus RD =0 cm is provided in table 5, but a breakdown of the range by study size and effect size is not presented but is available from the authors on request.

DISCUSSION

Principal findings

The elicitation exercise was likely to appeal to experts with polarised views of radical surgery and this was useful in getting representative opinion to inform priors.²⁶ It found that experts considered publication bias to be a possibility when assessing OS in the comparison of RD <1 cm versus RD =0 cm after PDS for EOC. This likelihood diminished considerably for the comparisons of suboptimal RD thresholds of >1 cm and >2 cm versus RD =0 cm, with most respondents (83.5%) believing it was not likely at all in comparison to RD >2 cm versus RD =0 cm. The

Table 5 Strength of responders' opinions as to likelihood of missing studies in RD <1 cm versus RD =0 cm and number of studies elicited

Strength of opinion of likelihood of missing studies	n	Estimated missing studies			Effect estimates*	
		Mean (SD)	Median (IQR)	Range	LogHR (SElogHR)	HR (95% CI)
'Not likely'	1	0	0	0	0 (0.25)†	1.0 (0.61 to 1.63)†
'Somewhat likely'	5	5.8 (2.4)	5 (5–5)	4–10	–0.098 (0.074)	0.91 (0.78 to 1.05)
'Quite likely'	8	17.8 (13.9)	12.5 (10–20)	7–50	–0.144 (0.054)	0.87 (0.78 to 0.96)
'Very/extremely likely'	4	37.5	40(25–50)	20–50	–0.078 (0.035)	0.92 (0.86 to 0.99)
All responders	18	17.8 (16.5)	10 (5–20)	0–50	–0.103 (0.066)	0.90 (0.79 to 1.03)

*Calculated using a simple weighted average of each responder.

†No studies were estimated from responder so for purposes of analysis and calculation of pooled estimate, one small and imprecise study was used.

RD, residual disease.

most striking finding was that experts were in large agreement about not needing to make any adjustments for publication bias in comparisons involving suboptimal cytoreduction versus complete resection, irrespective of role and surgical preference.

The average completion time of the elicitation exercise was quicker than the anticipated 30–60 min. This may have been due to some responders having an initial first look at the exercise before completing it during a later visit. This may help to explain the fastest completion time of 7.7 min. This hypothesis is consistent with how the exercise was designed, as we allowed up to 24 hours for completion following a first visit. In future work, we will consider a sensitivity analysis exploring the impact of excluding responses where completion times might be unrealistic.

Strengths and limitations

The elicitation exercise was designed in collaboration with senior gynae-oncologists. This is the main reason for the detailed level of explanation given, with visual examples, to ensure that potential respondents were clear about the tasks asked of them. This involved a trade-off between clarity of explanation and potentially dissuading some respondents from taking part. Our view was that getting data on a broader range of scenarios from a reduced number of respondents would be more valuable than getting data on a smaller number of scenarios from a greater number of respondents. This was not felt to be a major limitation as it has been argued that the opinions of only 4–16 experts are needed in expert elicitation exercises.^{28–31} The sample size achieved (n=18) was comfortably above this.

Part A of the elicitation exercise was based on an existing elicitation approach.³⁰ This part was used to identify areas where publication bias is of concern. Part B built on this by exploring the potential direction of bias. Therefore, in Part B of our elicitation exercise we collected information that would enable meta-analysis estimates to be adjusted for the impact of publication bias. The approach, while practical to use, relies on accurate survival estimates

being available as these are used to inform the study sizes. As noted above, it also requires that a sufficient number of experts provide an opinion (ie, 4–16).^{28–31}

Answers given by the experts to open-ended questions were prone to an 'extreme answer bias'. Therefore, we made the instructions that accompanied the elicitation exercise quite extensive. We discussed this in detail when we designed the exercise, and we feel more biases would be introduced if a ceiling of the number of estimated studies had been applied. Further work is planned to explore the impact of extreme responses on the conclusions drawn.

It is questionable as to whether the information gathered from any expert elicitation exercise can be considered a reliable estimate of relative effect. Therefore, its incorporation in a meta-analysis for adjustment may lead to 'more precise' estimates as shown by a CI but these may not be considered more reliable (that is we have gained precision but may have introduced another bias). The results shown in tables 1 and 3 appear to show variability in the answers given by the 18 respondents. Therefore, a series of sensitivity analyses would need to be conducted in order to test how robust the overall conclusions are to variations in the value of the priors used.

Implications for researchers and policy makers

Numerous recommendations have been put forth to help prevent publication bias in a systematic review, such as preregistration,³³ openness to negative or null findings by journal reviewers and editors,³⁴ use of preprint services to ameliorate the file-drawer problem,³⁵ and encouraging publication regardless of journal impact—which is often conflated as a metric of research quality.³⁶ These may offer a solution and minimise publication bias. However, they are not without issues. This leaves a need for methods that can instead allow us to explore and characterise the impact of publication bias. Our proposed method of expert elicitation can assist in this exploration.

The elicitation exercise provided results that may facilitate adjusting estimated effect sizes obtained with a meta-analysis for publication bias. Responders estimated that



data for substantial numbers of participants might be missing (eg, the estimate was over 3900 for the comparison of RD <1 cm vs 0 cm); this could have an impact on the results of meta-analyses. In particular, the responses from the elicitation exercise could be used to form Bayesian priors for a meta-analysis; specifically, the prior could be used to adjust the observed effect estimates obtained from the meta-analysis to explore the expected impact of publication bias. The 'educated guesses' from respondents are the only substantial source of information in this area that may facilitate such adjustment. The use of this method may be particularly important in situations like the one presented, where there is broad agreement that there is selective reporting and that there are unpublished studies that would provide 'non-significant' or 'negative' results. Should the estimates derived from the elicitation be used to adjust the meta-analysis comparing RD <1 cm, RD > 1 cm and RD >2 cm with RD =0 cm, we would expect that this would dilute the point estimate of the HR from any meta-analysis that suggested a benefit in OS for women whose tumour was cytoreduced to RD =0 cm. However, in this particular instance, there would be increased precision around the point estimate.

Within the online supplemental appendix 2, we outline one way in which such a prior could be formed from the collected data. In this approach, the weight given to each adjustment varies for the comparison of the different RD thresholds versus RD =0 cm. For example, respondents estimated more missing studies which included a greater number of participants for the comparison of RD <1 cm versus RD =0 cm. Consequently, the comparison of RD <1 cm versus RD =0 cm would have more influence in any adjustments made in a meta-analysis. Whereas, for the comparison of RD >2 cm versus RD =0 cm the estimates from the meta-analysis would be less affected as the consensus among responders of the exercise was that there was far less concern about publication bias. Furthermore, our illustrative approach gives each responder the same weight so that they contribute equally to the prior elicitation. However, we note that it would be possible to explore giving different groups a different weight. This might be relevant if we believed that different groups have different views on the nature and extent of missing data.

In meta-analyses assessing OS in suboptimal RD after PDS for advanced EOC, the evidence is relatively sparse, especially for RD thresholds >2 cm compared with RD =0 cm. For example, in our provisional scope of the results (necessary to facilitate Part A of the elicitation exercise), there was only one study that directly compared RD >2 cm versus RD =0 cm, and three studies where some indirect evidence relevant to this comparison was available. These four studies included only 478 women who contributed data for the comparison of RD >2 cm versus RD =0 cm. In this circumstance, the impact of prior expectations on the nature and extent of publication bias is likely to considerably affect the estimate. However, as evidence

accumulates, the weight given to a prior when making an adjustment to the meta-analysis result will be reduced.

Implications for clinicians

Publication bias can contribute to a false impression of the efficacy of a treatment effect or a prognostic factor within a body of literature.^{37 38} In the context of our expert elicitation exercise, publication bias appears to be most prone in the comparison of RD <1 cm and RD =0 cm. This may be due to the difficulty in knowing for sure that surgery has completely removed all tumour, as there still may be macroscopic disease. The a priori expectation is that this would bias the effect estimates in favour of near-optimal cytoreduction (RD <1 cm). The likelihood of publication bias comparing suboptimal cytoreduction >1 cm versus RD =0 cm was perceived by experts to be very low. If the literature is positively biased towards a certain conclusion, then meta-analyses will reflect that trend. Although there are assistive methods to help identify and expose publication bias such as funnel plots,¹⁸ they are by no means a full solution to the problem.

Research has shown that evidence from the literature is not the sole determining factor for clinical decision-making. Clinicians also have a preference for 'consensus-based decision-making' through relatively informal sources, such as their clinical colleagues and fellow academic experts. The opportunity to discuss and trade perspectives is treated as a valuable exchange to gather information and formulate one's judgement.^{39 40} Therefore, expert elicitation could be used to explore the impact of areas of uncertainty when developing clinical guideline recommendations.

Implications for future research

An extension to our work could be to build on the idea of using individual patient data (IPD) in meta-analyses,⁴¹ rather than using aggregate data. IPD can more easily incorporate a consistent selection of confounders to adjust for, which would reduce the impact of selective reporting of analyses and outcomes. An IPD analysis would also allow for comprehensive further exploration of confounders, which could include looking at possible interaction effects between confounders.⁴²

Additionally, it may not necessarily be missing studies that are the sole cause of publication bias; a systematic review can also be prone to the selective reporting of outcomes and analyses within published studies.^{9 43-45} This is an area that has been comprehensively critiqued and can be overcome to a large extent by conducting an IPD meta-analysis.⁴² Knowing that selective reporting is highly likely in the area under consideration, participants of the exercise potentially factored this into their elicitation estimates as, effectively, it equates to a missing study.

CONCLUSION

Previous evidence from meta-analyses suggests that complete cytoreduction of EOC is associated with

increased OS. However, our elicitation exercise of 18 experts also suggests that there is the potential for some concern about the nature and extent of publication bias in this area. The concerns are such that the unpublished evidence may substantially reduce or even remove the suggested OS benefit from complete cytoreduction compared with RD <1 cm. The results may raise awareness that a degree of scepticism is needed when reviewing studies comparing RD <1 cm versus RD =0 cm, especially when such evidence comes from non-randomised and sometimes post hoc analyses. Expert elicitation can be used to explore the impact of areas of uncertainty when developing clinical guideline recommendations. However, there is a strong belief among respondents that complete cytoreduction has an improved survival outcome compared with RD >1 cm and that publication bias is not related to that perception.

Author affiliations

¹Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

²Obstetrics and Gynaecology, Nottingham City Hospital, Nottingham, UK

³Pan-Birmingham Gynaecological Oncology Cancer Centre, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK

⁴Northern Gynaecological Oncology Centre, Queen Elizabeth Hospital, Gateshead, UK

Twitter Andrew Bryant @AndrewBryant82

Acknowledgements The authors thank Debbie Lewis from BGCS for excellent support with the conduct of the elicitation exercise. The authors also thank the BGCS organisation and survey committee for assistance in improving the content of the elicitation exercise and the opportunity to disseminate to members to complete. Judgements for each of the scenarios were based on the personal opinions of the sample of British Gynaecological Cancer Society (BGCS) members and reflected their own experience in this area. They do not necessary represent the views of BGCS as an organisation.

Contributors AB is leading author and conceptualised the methodology aspect of the research, drafted the paper, carried out the statistical analyses and is responsible for the overall content as guarantor. MG is senior statistician on the paper and reviewed methodology, analyses and critically inputted into sections of the paper. SH is a statistician who critically assessed sections of the paper and offered expertise in survey design. KG and AE provided clinical expertise and critical review as senior gynae-oncologists and researchers. EJ critically reviewed the paper and offered research experience in evidence synthesis. LV and DC are senior research academics who rigorously reviewed the methods and results and had input into the discussion. RN is a gynae-oncologist who conceptualised the clinical aspect of the research and rigorously inputted into the methods and discussion. All authors reviewed and approved the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This research involved human participants outside of a study or trial setting. The elicitation exercise did not require ethical approval because the elicitation exercise was sent to BGCS members and participation was optional. Information about any expert that participated in the elicitation exercise was kept confidential. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. An anonymised data set may be available on request and/or additional summary statistics provided.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Andrew Bryant <http://orcid.org/0000-0003-4351-8865>

REFERENCES

- Bryant A, Hiu S, Kunonga P. *Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery*. Cochrane Database of Systematic Reviews, 2021.
- Abbasi K. Compulsory registration of clinical trials. *BMJ* 2004;329:637–8.
- Moher D, Hopewell S, Schulz KF, et al. Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
- Moher D, Jones A, Lepage L, et al. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992–5.
- Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14.
- UKRI Medical Research Council. Mrc clinical trials review; 2019 [Accessed 01 Oct 2020].
- Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. *Evid Based Med* 2016;21:125–7.
- O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the Review Question and Developing Criteria for Including Studies. In: *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration, 2011. <https://handbook-5-1.cochrane.org/>
- Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration, 2011. <https://handbook-5-1.cochrane.org/>
- Schulz KF, Grimes DA, Altman DG, et al. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742–4.
- Burdett S, Stewart LA, Tierney JF. Publication bias and meta-analyses: a practical example. *Int J Technol Assess Health Care* 2003;19:129–34.
- Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration, 2011. <https://handbook-5-1.cochrane.org/>
- Ekmekci PE. An increasing problem in publication ethics: publication bias and editors' role in avoiding it. *Med Health Care Philos* 2017;20:171–8.
- Landewé RBM. Editorial: how publication bias may harm treatment guidelines. *Arthritis Rheumatol* 2014;66:2661–3.
- Egger M, Jüni P, Bartlett C, et al. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? empirical study. *Health Technol Assess* 2003;7:1–82.
- Hopewell S, McDonald S, Clarke M, et al. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev* 2007;2:MR000010.
- Mallett S, Hopewell S, Clarke M. Grey literature in systematic reviews: The first 1000 Cochrane systematic reviews. In: *47th*



- Symposium on systematic reviews: pushing the boundaries*. Oxford, UK, 2002.
- 18 Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;54:1046–55.
 - 19 Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455–63.
 - 20 Duval S, Tweedie R. Practical estimates of the effect of publication bias in meta-analysis. *Australasian Epidemiologist* 1998;5:14–17.
 - 21 Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979;86:638–41.
 - 22 Copas JB, Malley PF. A robust P-value for treatment effect in meta-analysis with publication bias. *Stat Med* 2008;27:4267–78.
 - 23 Sutton AJ, Song F, Gilbody SM, et al. Modelling publication bias in meta-analysis: a review. *Stat Methods Med Res* 2000;9:421–45.
 - 24 Egger M, Smith GD, Altman DG, eds. *Systematic reviews in health care: Meta-analysis in context*. London, UK: BMJ Publishing Group, 2001.
 - 25 Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. Chichester, UK: John Wiley & Sons, 2004.
 - 26 Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res* 2001;10:277–303.
 - 27 Sutton AJ, Abrams KR, Jones DR. *Methods for meta-analysis in medical research*. Chichester, UK: John Wiley & Sons, 2000.
 - 28 Colson AR, Cooke RM. Expert elicitation: using the classical model to validate experts' judgments. *Rev Environ Econ Policy* 2018;12:113–32.
 - 29 Cooke RM, Goossens LJH. Procedures guide for structured expert judgment. European Commission; 2000.
 - 30 Mavridis D, Welton NJ, Sutton A, et al. A selection model for accounting for publication bias in a full network meta-analysis. *Stat Med* 2014;33:5399–412.
 - 31 Wilson ECF, Usher-Smith JA, Emery J, et al. Expert elicitation of multinomial probabilities for decision-analytic modeling: an application to rates of disease progression in undiagnosed and untreated melanoma. *Value Health* 2018;21:669–76.
 - 32 StataCorp. *Stata statistical software: release 15*. College Station, TX: StataCorp LLC, 2017.
 - 33 Nosek BA, Ebersole CR, DeHaven AC, et al. The preregistration revolution. *Proc Natl Acad Sci U S A* 2018;115:2600–6.
 - 34 Bernalov A, Steckler T, Skolnick P. Be positive about negatives—recommendations for the publication of negative (or null) results. *Eur Neuropsychopharmacol* 2019;29:1312–20.
 - 35 Verma IM. Preprint servers facilitate scientific discourse. *Proc Natl Acad Sci U S A* 2017;114:12630.
 - 36 Bohannon J. Hate Journal impact factors? new study gives you one more reason. *Science* 2016.
 - 37 Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. *Mayo Clinic Proceedings* 2000;75:1284–8.
 - 38 Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;6:k4597.
 - 39 Gupta DM, Boland RJ, Aron DC. The physician's experience of changing clinical practice: a struggle to unlearn. *Implementation Science* 2017;12:28.
 - 40 Kristensen N, Nymann C, Konradsen H. Implementing research results in clinical practice— the experiences of healthcare professionals. *BMC Health Serv Res* 2015;16:48.
 - 41 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
 - 42 Riley RD, Debray TPA, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: statistical recommendations for conduct and planning. *Stat Med* 2020;39:2115–37.
 - 43 Dwan K, Altman DG, Clarke M, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* 2014;11:e1001666.
 - 44 Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 2005;24:1547–61.
 - 45 Williamson PR, Gamble C, Altman DG, et al. Outcome selection bias in meta-analysis. *Stat Methods Med Res* 2005;14:515–24.
 - 46 Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998;17:2815–34.