SIMETRICA Jacobs

DCMS Rapid Evidence Assessment: Culture and Heritage Valuation Studies - Technical Report

October 30, 2020

Lawton, R, Fujiwara, D, Arber, M, Maguire, H, Malde, J, O'Donovan, P, Lyons, A., Atkinson, G.

The Shepherds Building,
Richmond Way, W14 0EH
0203 883 9249
information@simetrica.co.uk
simetrica.co.uk

Table of Contents

Executive Summary

The culture and heritage sectors have a significant impact on society in the UK, particularly given our rich history and efforts to preserve cultural and heritage assets. Whilst there is a wealth of literature aiming to value different aspects of culture, art and heritage, there exists a lack of a formalised approach. DCMS are launching a formal Culture & Heritage Capital approach (similar to the natural capital approach) to provide standard methods to value stocks and flows and would contribute to informing funding decisions. To learn more about the DCMS's approach, aims and future research see "Valuing Culture & Heritage Capital: A framework towards informing decision making" published alongside this report. Before more formal work begins on a capital approach, the Department for Digital Culture Media and Sport (DCMS) is seeking **better knowledge of valuation techniques and reliable values for a range of culture and heritage assets to inform government spending rounds and business investment decisions.**

The aim of a **Rapid Evidence Assessment (REA)** is to provide a balanced systematic assessment of what is known about a policy issue and what gaps may remain, whilst making concessions to the breadth or depth of the process due to the short time period available[1]. This assessment method is therefore an excellent way of determining what is known in the literature about a given topic and what needs further research. Priority areas for future research highlighted by the REA have been listed in the CHC framework published alongside this document. The REA is also designed to help direct ongoing research in the cultural sector and will inform the Culture & Heritage Capital programme currently in development at DCMS.

This REA was conducted to **assess the current state of the literature valuing cultural and heritage assets to determine those that align with best practice methods and identify weaknesses/gaps.** The results are presented within an Evidence Bank of economic values. The Evidence Bank includes valuation details, such as estimated monetary values for assets, a grading of the quality of each study, the article details, and an overview of each valuation method used.

The focus of this REA is on **studies that employ economic approaches for monetary valuation of culture and heritage asset**s. It is important to make clear at the outset a distinction between different concepts of value.

Culture and Heritage physical assets were defined as:

- **Archaeological assets:** castles, ruins

[1] HM Treasury 2011

- **Art engagement**: public arts, street art, festivals, libraries digital and heritage archives, and valuing regular participation or engagement in the arts more broadly
- **Built heritage**: towns, cities, businesses, cinemas, plaques, built, and buildings
- **Cultural institution**: art galleries, museums, concert or town halls, music venue or bandstands or amphitheatres, theatres or playhouses or opera houses
- **Digital assets**: including public service broadcasting
- **Industrial heritage**: transport, roads, rail, bridges, canals, mines, quarries, warehouses, mills, factories, waterways, ports, docks, harbours, and aqueducts
- **Historic amenities**: monuments, sculptures, statues, structures, gardens, parks, and landscapes
- **Protected area**: areas, sites, places, and spaces
- **Religious asset**: cathedrals, churches, chapels, mosques, temples, synagogues, monasteries, shrines.

This review particularly focuses on valuations of assets that are public goods and so are free to access (e.g., museums, historical cityscapes). Such assets rely on non-market valuation, which encompasses a much wider range of impacts covering society more widely, by looking at the value that individuals place on a good or service. Non-market valuation techniques allow researchers to value cultural goods and services that are not traded in markets. The main non-market valuation techniques are:

Revealed Preference (RP) methods applied to goods and services that result in observable changes in behaviour in indirect markets (e.g., the value of built heritage may be revealed indirectly in housing markets across regions where the level or quality of provision of built heritage differs).

Stated Preference (SP) methods applied to goods and services that do not result in observable changes in market behaviour but are amenable to direct monetisation (e.g., willingness to pay to access a hypothetical entry fee to access a cultural institution that is currently free to the public).

Wellbeing Valuation (WV) methods applied to goods and services that do not result in observable changes in market behaviour and are difficult to monetise directly, but may have measurable effects on individual wellbeing measures and so can be monetised indirectly (e.g., regular engagement with culture and heritage).

Benefit, or Value, Transfer (BT), is the method of transferring values from one site to another. Values can be obtained from the literature using source studies (or source sites) rather than costly primary data

collection. Source site values can include use and non-use values obtained from SP and RP studies.

This REA was informed by consultation with Professor Giles Atkinson of the London School of Economics (LSE), and stakeholders from across Government. The methodology for this report was formulated with transparency in mind, so the process and findings of the report can be replicated by other researchers across all sectors. To maximise the cost-effectiveness of our work, we employed data science techniques (i.e., text-mining) to efficiently search for the most relevant literature and verify the robustness of our search strategies. Text-mining helped to speed up the search for relevant papers, data extraction, and data integration processes through topic identification.

It has been nearly 20 years since Noonan et al.'s (2002) notable meta-analysis of valuation studies in the cultural sector, with a large amount of new empirical evidence produced in the culture and heritage sector since then. There is growing real-world demand from policymakers and practitioners to monetise the welfare impacts of cultural policies, and to quantify culture and heritage goods and services in a way that aligns with cost benefit analysis and business cases in other sectors. There has also been an advancement in guidelines for good practice valuation studies.[2] This report is therefore a timely opportunity to take stock of the field, draw conclusions, and identify common standards of practice to help inform designing best-practice valuation studies, structured interpretation of evidence and identification of robust examples. This work can also aid in triangulating the relative values obtained through different valuation methods (e.g., RP and SP), providing the potential to start estimating and suggesting corrections for these known differences in the scale of values obtained by different methods.

1.1 Quality Criteria Rating

Using a novel approach, we used **data science techniques** to collect 171 relevant, academic papers and grey literature reports (Government reports and private consultancy publications) on valuation of culture and heritage assets. which were screened first by machine learning code and second by researcher investigation. Papers were then screened for quality by researchers for more complex criteria, by grading the quality of the research papers based on the **Quality Grading Criteria**. The REA found that over two-thirds (116) of culture and heritage values were rated as medium to high quality by the Quality Grading Criteria. As shown in Figures 0-1 and 0-2, the number of values and distribution by quality rating varies by valuation methodology. SP values, which are predominantly CV, are by far the most common. Most methodologies have a clear mix in the quality of values.

[2] Bateman et al. 2002; R. J. Johnston et al. 2017; G Atkinson et al. 2018

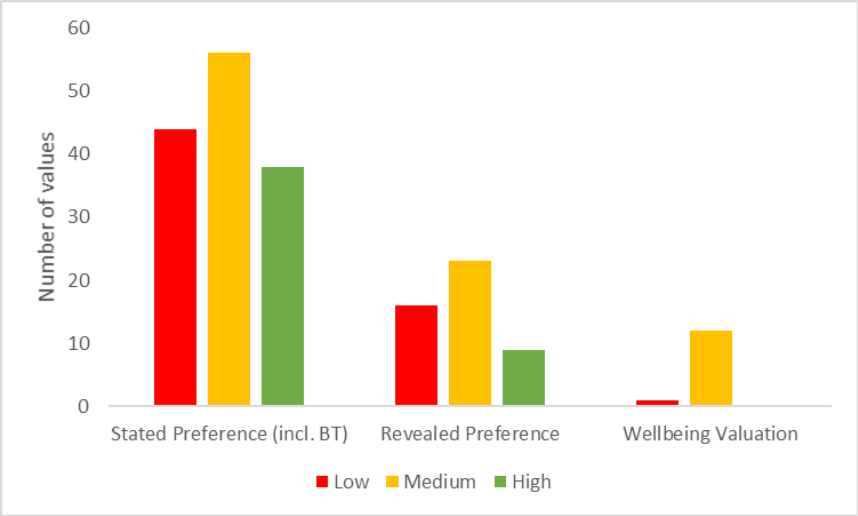


Figure 0-1 Distribution of values by high-level valuation methodology and quality rating

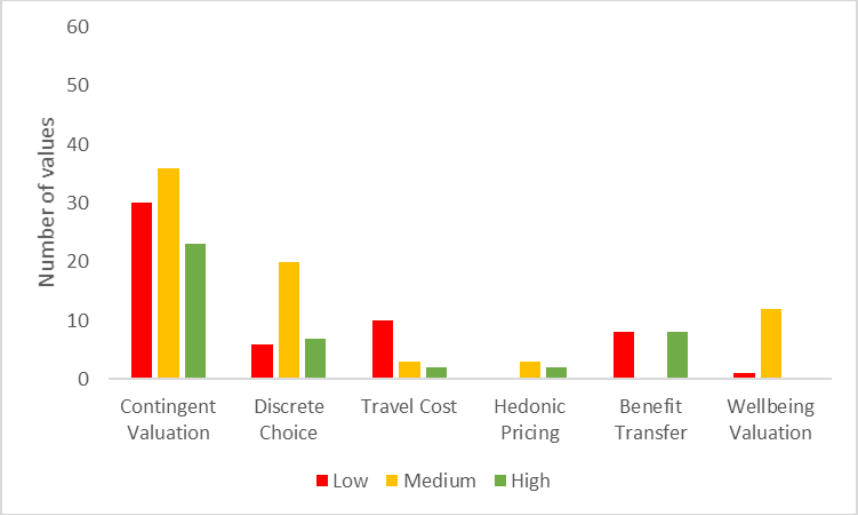


Figure 0-2 Distribution of values by valuation methodology and quality rating

The Quality Grading Criteria developed for this REA has never been attempted before in the cultural or environmental sector, and its synthesis of best practice guidance[3] and direct researcher experience provides a quality RAG rating for all of the values in the Evidence bank, representing a considerable contribution to the sector.

1.2 Key findings

The REA results suggest that **established non-market valuation methods, using SP and RP methods, are more prevalent in the academic literature than the more recent and novel methods like WV and BT.** In the grey literature,

[3] G Atkinson et al. 2018; Bateman et al. 2002; R. J. Johnston et al. 2017

Contingent Valuation (CV) SP studies (either individual studies or as part of a BT) are the most common, but there is also a rising number of WV studies. We hypothesize that this may be driven partly by policy need, since CV and WV provide valuations for the cultural heritage asset as a whole, rather than attributes of the asset (as in Discrete Choice Experiments (DCE)) and this is more relevant for government business cases and funding bids.

1.3 Thematic split (type of culture or heritage asset)

Within both the academic and grey literature the **highest number of culture and heritage values were found for cultural institutions**, followed by archaeological assets, historical amenities and built heritage. Cultural institutions were also found to have the highest number of medium to high quality valuation ratings in the literature (see *Figure 0-3*).

The **lowest number of culture and heritage values** in both academic and grey literature were found for **industrial heritage, protected areas and religious assets.**

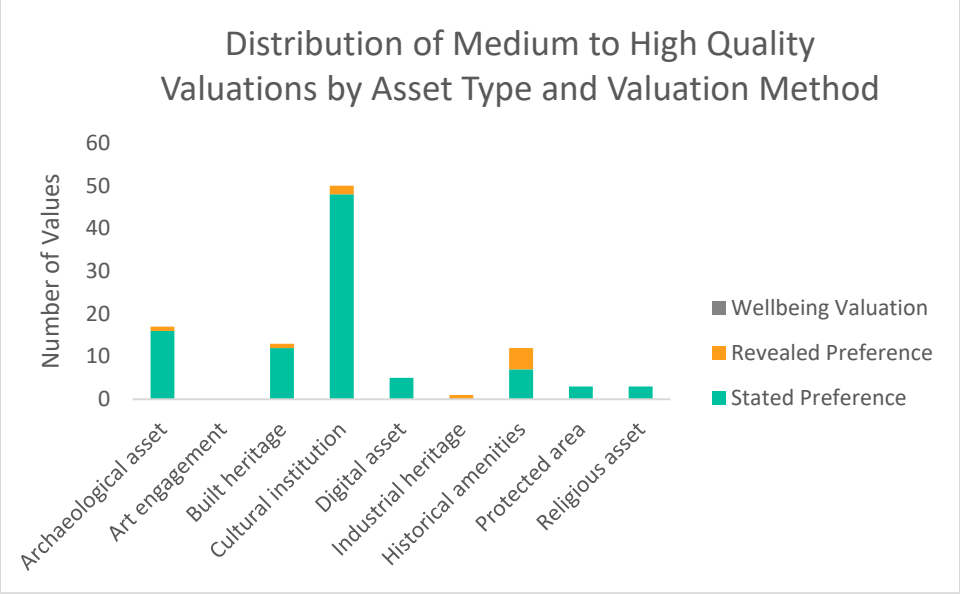


Figure 0-3 Distribution of Medium to high Quality Valuations by Asset Type and Valuation Method

Overall, there is evidence of a sometimes **patchy empirical record** that is well represented for certain types of tangible culture and heritage asset, such as **museums/galleries/theatres and libraries and built heritage sites.** Less well represented in the valuation literature are **digital services, industrial heritage, and religious assets.** There is also a **narrow range of non-market valuation methods** used, in particular CV and DCE, although there may be some evidence of **evolving trends** in the enhanced use of WV, which uses observational data estimate the value that individuals gain from engagement or participation with culture and heritage, and BT, which involves

transposing 'primary' research valuation estimates from one site to another in a fast and cost-effective way as it means the valuation estimates can be used in other contexts.

1.4 Research gaps

Very few non-market values were identified for industrial heritage or religious assets. This suggests that there is an evidence gap for DCE valuation of urban built heritage (excluding archaeological assets) which can be filled by commissioning research in this area in the future.

Digital cultural heritage assets and services (excluding Public Service Broadcasting) have not been valued using non-market valuation methods in the academic literature to date, which represents an important research gap going forward. The **absence of values for other types of digital assets** is surprising, given the recent emergence of online portals for digital cultural services in many countries (such as virtual tours of art galleries) and the digitisation of cultural archives (such as the British Film Institute Unlocking Film Heritage project[4]). This shift to digital services has not yet been reflected in the non-market valuation literature, either in the academic or the more responsive grey literature fields. We anticipate that this may experience a step change in the coming decade, especially in response to the 2020 COVID-19 pandemic which led to the closure of many cultural institutions for many months and an acceleration of their digital offer.

The small number of hedonic values in the culture and heritage sector indicates that there is an opportunity gap in valuation research to exploit enriched real estate data to value those types of culture and heritage assets which are amenable to Hedonic Price (HP) methods, notably heritage buildings and protected areas within urban zones. Increased policy efforts to **make these proprietary real estate data datasets open-source for culture and heritage research** could help to overcome this gap in the future. Where HP models have been applied to culture and heritage assets, the REA suggests that datasets available for hedonic analysis are strong across multiple jurisdictions, and we would expect this to continue into the future with the opening up of online land registry and house market data. This has resulted in a healthy set of hedonic values for those cultural heritage asset types which are amenable to this method (i.e., those assets which are expected to affect local house prices).

Similarly, there may be **opportunities to use the Travel Cost Method (TCM) in a more systematic way to value audience engagement with culture and heritage sites using existing time-use surveys collected at the national level.**[5] There may also be existing models from other sectors that could be followed to

[4] https://www.bfi.org.uk/britain-on-film/unlocking-film-heritage
[5] https://data.gov.uk/dataset/ba226a1a-de1e-4f9e-80e6-835e3518764e/time-use-surveys-and-the-measurement-of-national-well-being

provide Travel Cost estimates of the value of visits to heritage or religious sites in the UK. For instance, Defra produced the Outdoor Recreation Valuation tool (ORVAL) to understand how people value the country's natural environments and greenspaces using the Monitoring of Engagement in Natural Environment Survey. It may be possible to combine this model with emerging data visits to cultural institutions, heritage sites and religious buildings. Even accepting that the TCM provides a partial value for the value of a trip (rather than the total use and non-use values of the site), it would still be useful to know the value of cultural resources even in those narrower terms.

An additional gap in the research is that we found no valuation studies that apply **welfare weighting** to the values estimated for culture and heritage assets. Welfare weighting, as recommended by HM Treasury Green Book, permits using **distributional weights to adjust for diminishing marginal utility of income**. We found **no instances in which welfare weighting was pursued in the culture and heritage REA**. Consequently, non-market valuation in the culture and heritage sector does not account for how inability to pay might constrain stated willingness to pay values, possibly leading to inflated willingness to pay values for sites frequented by higher socioeconomic users compared to sites which are frequented less by such users. A distorted funding decision may arise as a potential result of this, as sites with higher socioeconomic visitors, and thereby greater aggregated value, might access more funding than sites without this audience. We recommend that future applications of non-market valuation methods consider the application of welfare weighting in instances where there may be significant differences between the values held by gainers and losers in the population.

2 Introduction

The culture and heritage sectors have a significant impact on society in the UK, particularly given our rich history and efforts to preserve culture and heritage assets. Whilst there is a wealth of literature aiming to value different aspects of culture, art and heritage, there exists a lack of a formalised approach. A formal Culture & Heritage Capital approach (similar to the natural capital approach) would provide standard methods to value stocks and flows and would contribute to informing decision-making on public funding.

In order to gain a better understanding of valuation techniques and obtain reliable values across the arts, museums and heritage sectors, DCMS commissioned a Rapid Evidence Assessment (REA) of existing valuation studies.

The aim of a Rapid Evidence Assessment is to provide a balanced systematic assessment of what is known about a policy issue and what gaps may remain whilst making concessions to the breadth or depth of the process due to the

short time period available.[6] This assessment method is therefore an excellent way of determining what is known in the literature about a given topic, and what needs further research. The REA is designed is to help to direct ongoing research in the cultural sector and will inform the Culture & Heritage Capital programme currently in development at the Department of Digital, Culture, Media and Sport (DCMS).

This REA was conducted to assess the current state of the literature valuing culture and heritage assets to determine those that align with best practice methods and identify weaknesses/gaps. The results are presented within an Evidence Bank of economic values. The Evidence Bank includes valuation details, such as estimated monetary values for assets, a grading of the quality of each study, the article details, and the overview of each valuation method used.

The focus of this REA is on studies that employ economic approaches for monetary valuation of culture and heritage assets. It is important to make clear at the outset a distinction between different concepts of value.

As formally defined, economic value refers to the impact of changes in a good or service on human welfare (or wellbeing), measured in monetary terms. In economics, we measure the value of a good, service or outcome in terms of its impact on human welfare, expressed either as a compensating welfare measure (the amount of money, to be paid or received, that will leave the agent in their initial welfare position following a change in the status quo) or as an equivalent welfare measure (the amount of money, to be paid or received, that will leave the agent in their subsequent welfare position in absence of a change from the status quo), where a 'change' is in the form of price changes or changes to the quantity or quality of the good, service or outcome. For goods that are traded in a market, a value can be inferred from the good's market price. For goods that are not traded in a market, so called 'non-market goods', alternative methods for valuation are required. Figure 1-1 shows how non-market value can be broken down into use and non-use values (note that not all non-market valuation methods are capable of eliciting the full range of these values, as detailed in Appendix Table 6-1).

[6] HM Treasury 2011

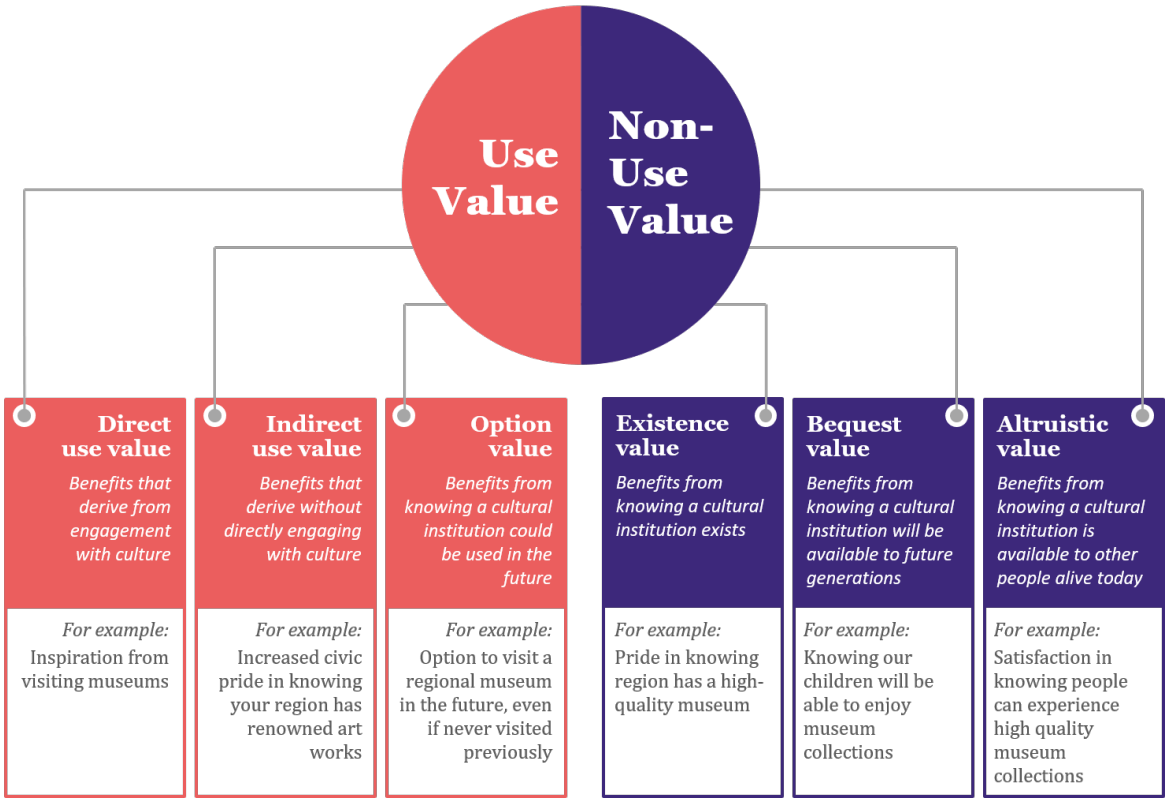


Figure 1-1 Non-market value: Use and non-use values within a Total Economic Value framework

This review particularly focuses on valuations of assets that are public goods; free to access (e.g., museums, historical cityscapes). Such assets rely on non-market valuation, which encompasses a much wider range of impacts covering society more widely, by looking at the value that individuals place on a good or service. Non-market valuation techniques allow researchers to value cultural goods and services that are not traded in markets:[7]

Revealed Preference methods applied to goods and services that result in observable changes in behaviour in indirect markets (e.g., the value of built heritage may be revealed indirectly in housing markets across regions where the level or quality of provision of built heritage differs). Although typically based on property markets in culture and heritage valuation, in principle, hedonic wage studies could also plausibly detect wage discount for working in culturally rich area. In practice, the statistical difficulties of identifying this effect might be a reason why only Hedonic House Price studies were found in the REA.

Stated Preference methods applied to goods and services that do not result in observable changes in market behaviour but are amenable to direct

[7] Other methods, such as the Delphi Method, were beyond the scope of the current review as this method is not outlined in the HM Treasury Green Book.

monetisation (e.g., willingness to pay to access a hypothetical entry fee to access a culture or heritage institution that is currently free to the public).

Wellbeing Valuation methods applied to goods and services that do not result in observable changes in market behaviour and are difficult to monetise: (e.g., regular engagement with culture and heritage).

Benefit, or Value, Transfer, is the method of transferring values from one site to another. Values can be obtained from the literature using source studies (or source sites) rather than costly primary data collection. Source site values can include use and non-use values obtained from SP and RP studies. We note that BT is not a non-market valuation method in its own right, as it relies on an empirical record to transfer from. Consequently, there is a distinction to make between primary and secondary valuation methods. However, given the policy relevance of the BT approach, it is given a separate section within the REA.

Appendix Table 6-1 summarises the key characteristics of each of the main techniques covered in these methods.

Our findings from this REA are presented in this technical report and accompanying Evidence Bank. While we have selected the most robust results and equivalised these values for our evidence bank, we do not recommend directly transferring these values into case studies. Caution should be taken into consideration when interpreting these values.

3 Literature review: Previous reviews of culture and heritage valuation studies

Empirical research eliciting economic values or benefits associated with access, preservation or restoration of cultural assets dates back to the 1980s when the first valuation studies in the field were conducted, focusing on the arts, theatre, historical sites, museums, galleries, libraries and broadcasting.[8] Since then, many studies in the cultural sector have been conducted worldwide investigating a variety of benefits, both tangible and intangible. In the cultural and heritage field, tangible benefits refer to the positive effects experienced from, for instance, the attraction of tourists and residents to a place. The more abstract intangible benefits arise from the increased desire to preserve cultural meaning or significance of a place.

[8] Noonan 2003

Over the past three decades there has been growing appreciation that the value of culture and heritage can be captured through preference-based empirical techniques, such as SP and RP methods. This is in contrast to the formal designations and valuations of significance based on expert knowledge (e.g., Delphi valuation method)[9]. Preference-based methods aim to elicit visitor and/or non-visitor willingness to pay values. But in recent years, studies using WV have also begun to appear. In the academic literature, detailed reviews of SP heritage valuation studies include Noonan[10], Pearce et al.[11], and Provins et al.[12] However, none of these reviews include WV studies (see Appendix 6.1 for full review of previous meta-reviews).

In the grey literature (defined in this report as Government reports and private consultancy publications which are not traditionally reviewed by a third party), the Department for Environment, Food and Rural Affairs (Defra, 2020) published a comprehensive methodological literature review across ecosystem service areas, including natural landscapes and tourism.[13] This Enabling Natural Capital Approach (ENCA) guidance provides physical and economic valuation evidence by service and asset category, overarching guidance for the natural capital field, and featured tools and case studies.

As a result, there is a gap for a review of the culture and heritage valuation literature that includes other non-market valuation methods, such as RP (HP and TCM) and the recently emerged field of WV. Many of the previous reviews focused only on a small number of culture and heritage categories, included only academic literature, or limited their focus on one or two economic valuation methods. By excluding the grey literature, these reviews have excluded empirical evidence which, while not peer-reviewed, may nonetheless provide important evidence of cultural values. The field also lacks a review that includes an assessment of the relative quality of the papers reviewed, provides a summary of the trends, and acknowledges the current state of the valuation research.

Given that it has been nearly 20 years since Noonan et al.'s (2002) notable meta-analysis of valuation studies in the cultural sector with a large amount of new empirical evidence produced in the culture and heritage sector since then. There is growing real-world demand from policymakers and practitioners to monetise the welfare impacts of cultural policies, and to quantify culture and heritage goods and services in a way that aligns with cost benefit analysis and business cases in other sectors. There has also been an advancement in guidelines for good practice valuation studies.[14] This report is therefore a timely opportunity to take stock of the field, draw

[9] Crossick and Kaszynska 2016; Noonan 2003; Throsby 2003; K.G. Willis 2014; Provins et al. 2008
[10] Noonan 2003
[11] Pearce and O'zdemiroglu 2002
[12] Provins et al. 2008
[13] https://www.gov.uk/guidance/enabling-a-natural-capital-approach-enca
[14] Bateman et al. 2002; R. J. Johnston et al. 2017; G Atkinson et al. 2018

conclusions, and identify common standards of practice to help inform designing best-practice valuation studies, structured interpretation of evidence and identification of robust examples.

4 Data and methodology

The REA followed a standard four-stage process outlined below in Figure 3-1 The REA process in four stages. Stages 1–4 are described in detail in the Technical Appendix 6.2. Possible limitations to the REA process and the methods employed are discussed in the appendix and concluding section of this report.

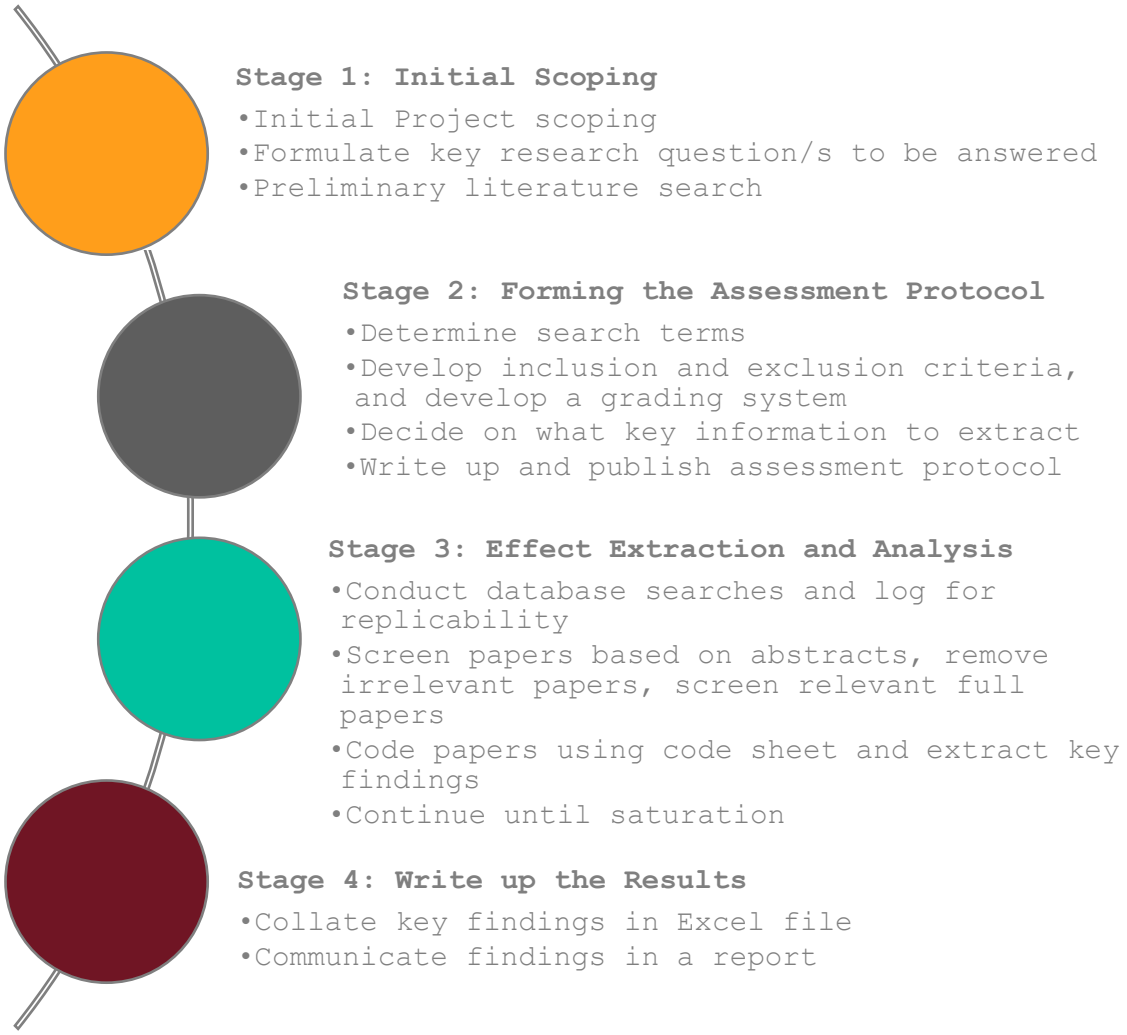


Figure 3-1 The REA process in four stages

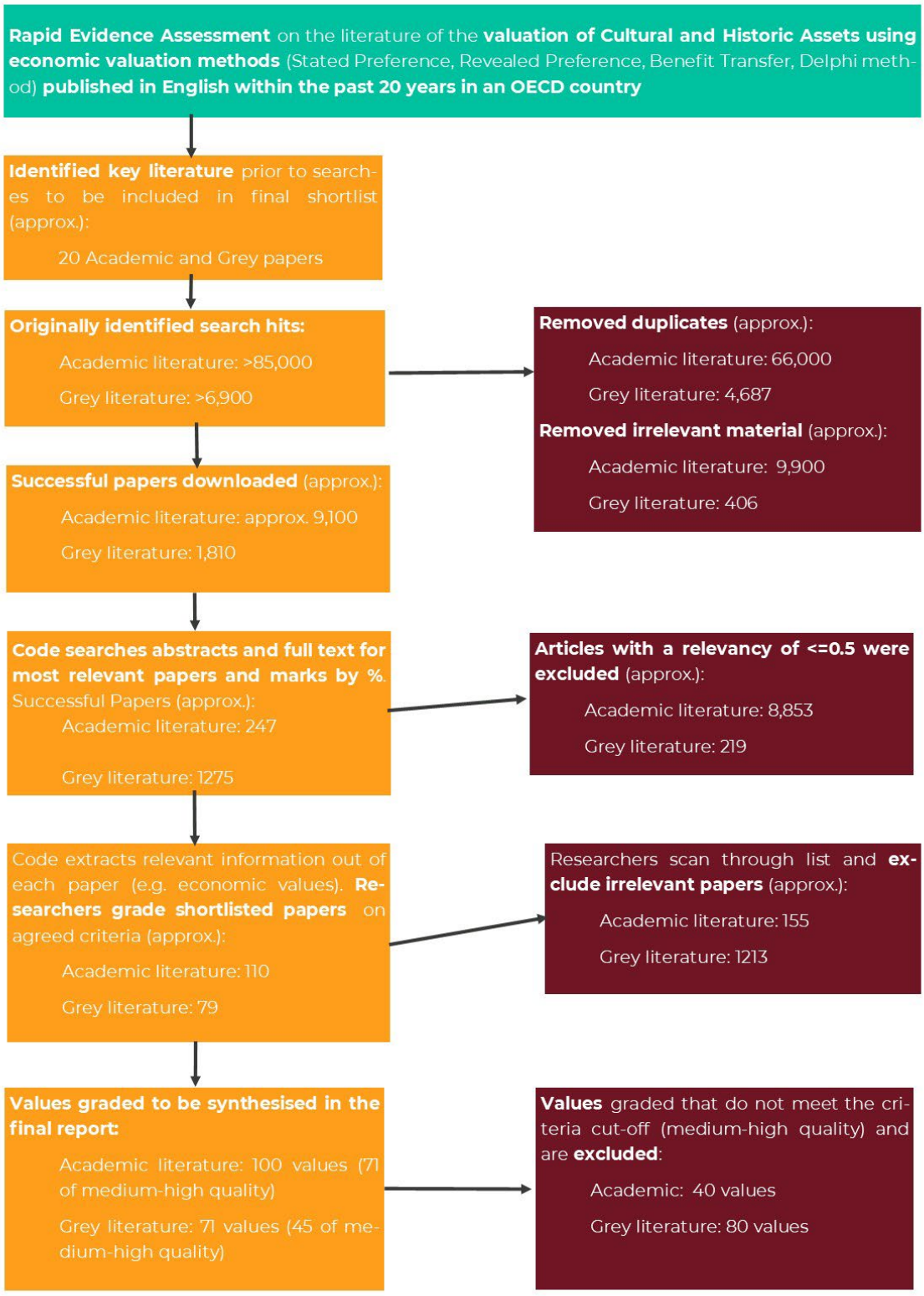


Figure 3-2 Project flowchart

5 Results

5.1 Stage 4: REA Results

5.1.1 Overview of Literature

The culture and heritage values in the Evidence Bank are reported as 2020 UK average £ value per person adjusted for purchasing power parity. A detailed explanation of the steps taken to convert all values to an equivalent per person GBP value is provided in Appendix 6.2.5. All values elicited for a household are equivalised by dividing by the average number of individuals living in a household in the country in which the study was based.

In the sections below we summarise the key findings of the Evidence Bank. All tables provide total numbers of values identified through academic literature, grey literature, and the complete dataset of values found across the REA.

Table 4-1 shows the total of 171 non-market valuations of cultural and heritage goods and services identified in the culture and heritage academic and grey literature since 2000, split by non-market valuation methodology. The values are broken down by total number of values and the number of values assessed to be of good quality (i.e., rated medium or high quality overall in the Quality Grading Criteria). This Quality Criteria RAG rating was developed in collaboration with stakeholders prior to the searches being conducted with specific exclusion criteria based on the chosen economic valuation methods. This criterion determined rigorous standards for each evaluation method. A full set of criteria for low, medium, or high-quality rating is provided for each method in the Evidence Bank and summarised in Appendix 6.2.4. The REA found that over two-thirds (116) of culture and heritage values were rated as medium to high quality by the Quality Grading Criteria.

All valid papers are included in the Evidence bank, regardless of Quality Criteria RAG rating. Although we have included studies with low RAG ratings in the evidence bank, we provide strong caveats against following these methods and using values from these studies. We only provide detailed reports of studies which were assessed to have an overall medium to high quality according to the Quality Grading Criteria.

It should be noted that we talk here of 'number of values' rather than number of studies, since one study can provide multiple values for a single asset or multiple assets.[15] Wherever possible we use the researcher analysis to identify the author's preferred estimate or that which we consider to be the

15 As noted above, it was out of scope of the REA to test for correlation of WTP values within single studies. However, this would typically only be necessary for the purposes of statistical meta-analysis, which is not pursued here.

most robust. For example, some SP studies report WTP values with and without those not willing to pay in principle for sensitivity analysis (i.e., those willing to pay £0) but the appropriate average estimate should always include those not willing to pay in principle.

Table 4-1 Non-market valuations of cultural and heritage goods and services 2000-2019, by methodology

Method	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Stated Preference: Contingent Valuation	52	37	37	22	89	59
Stated Preference: Discrete Choice	29	25	4	2	33	27
Revealed Preference: Travel Cost	13	5	2	0	15	5
Revealed Preference: Hedonic (House) Pricing	4	4	1	1	5	5
Benefit Transfer: Stated Preference: Contingent Valuation	0	0	15	8	15	8
Benefit Transfer: Stated Preference: Discrete Choice	0	0	0	0	0	0
Benefit Transfer (other)	1	0	0	0	1	0
Wellbeing Valuation	1	0	12	12	13	12
Total	100	71	71	45	171	116

5.1.2 Academic Literature Values

The REA included 100 peer-reviewed academic values, of which 71 were rated by our researchers as good (medium to high) quality. The most common type of non-market methodology for valuing tangible culture and heritage assets is SP, both CV (52 values, of which 37 rated medium-high quality) and DCE (29 values, of which 25 rated medium-high quality).

RP values were the next most common method, with 13 Travel Cost values and 4 hedonic values. However, most Travel Cost values were of low methodological quality. For instance, being based on single-site rather than multi-site visits, thereby failing to account for the availability of alternative and complementary activities, where travel to the location of a cultural heritage site is undertaken for more than one purpose.[16] This left only 5 Travel Cost values of medium-high quality. HP values were the most robust, with all 4 values obtained using this method being assessed as medium-high quality according to the Quality Grading Criteria.

Only one academic BT study was found, which used a mixture of SP and RP methods; however this study was graded as a low quality due to using 3 source sites, absence of transfer error testing, heterogeneity in valuation approaches, and only providing an aggregated value estimate.

Likewise, WV was less common in the academic search, with only 1 value which was rated low quality due to poor sample size and quality of modelling.

5.1.3 Grey Literature Values

The REA included 71 values from grey literature sources, of which 45 were rated as good (medium to high) quality. Again, the most common type of non-market methodology for valuing tangible culture and heritage assets is SP CV, both in standalone studies (37 values, of which 22 are rated medium-high quality) and through BT (15 values, of which 8 rated medium-high quality). We found no examples of BT using any other non-market valuation method.

DCE was less common in the grey literature than academic studies (4 values, of which 2 rated medium-high quality).

WV is much more common in the grey literature than academic literature (12 values, all rated medium-high quality). This makes WV of culture and heritage assets more common than RP Travel Cost (2 values, of which 0 were rated medium-high quality) and Hedonic methods combined (1 value rated medium-high quality) in the grey literature.

All wellbeing values were estimated since 2013 (compared to 2000 for SP and 2004 for RP studies) indicating that this method has grown in prominence in the grey literature but has perhaps been delayed in the academic literature, which requires longer timeframes between submission and publication. We may therefore expect that in the next 5-10 years, a larger number of peer-reviewed wellbeing values will be published. We note that the majority of

[16] G Atkinson et al. 2018

these studies value regular engagement with cultural institutions, rather than single visits or the wider value of heritage to the local environment..

5.1.4 Thematic split

The culture and heritage values obtained through the REA cover the topic areas outlined in Table 4-2, which we group under the following umbrella categories for ease of reference. We also report the number of medium-high quality values for each asset type (see Figure 4-1).

- Archaeological asset: 16 academic peer-reviewed values and 1 grey literature value of medium-high quality
- Art engagement (valuing regular participation or engagement in the arts more broadly): No academic peer-reviewed values, and 3 grey literature values of medium-high quality
- Built heritage: 7 academic peer-reviewed values and 6 grey literature values of medium-high quality
- Cultural institution: 35 academic peer-reviewed values and 19 grey literature values of medium-high quality
- Digital asset: 1 academic peer-reviewed value and 4 grey literature values of medium-high quality
- Industrial heritage: no academic peer-reviewed values and 1 grey literature value of medium-high quality
- Historical amenities: 10 academic peer-reviewed values and 7 grey literature values of medium-high quality
- Protected area: 1 academic peer-reviewed values and 2 grey literature values of medium-high quality
- Religious asset: 1 academic peer-reviewed values and 2 grey literature values of medium-high quality

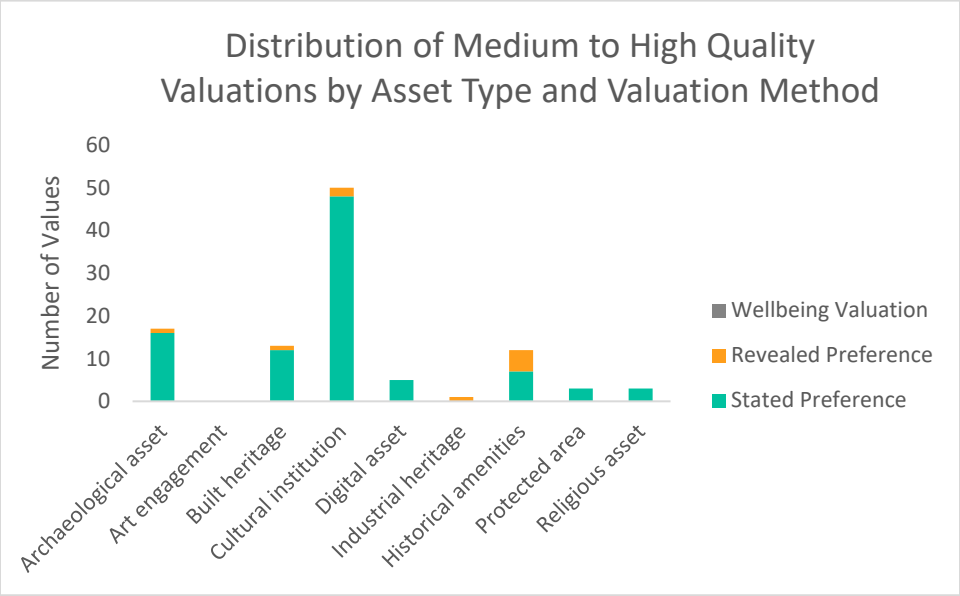


Figure 4-1 Distribution of Medium to high Quality Valuations by Asset Type and Valuation Method

In sum, within both the academic and grey literature the highest number of culture and heritage values (rated medium-high quality) were found for cultural institutions (35 academic and 19 grey literature). The academic literature had more values for archaeological asset (16 academic values compared to only 1 grey literature value). Historical amenities and built heritage were well covered in both academic (10 and 12 values respectively) and grey literature (7 and 7 values respectively).

The lowest number of culture and heritage values in both academic and grey literature were found for industrial heritage, protected areas and religious assets.

A higher number of digital assets were valued within the grey literature (n=4) than academic literature (n=1).

Overall, the absence of values for other types of digital assets is surprising, given the recent emergence of online portals for digital cultural services in many countries (such as virtual tours of art galleries and the digitation of cultural archives (such as the British Film Institute Unlocking Film Heritage project[17]), this shift to digital services has not yet been reflected in the non-market valuation literature, either in the academic or the more responsive grey literature fields. We anticipate that this may experience a step change in the coming decade, especially in response to the

[17] https://www.bfi.org.uk/britain-on-film/unlocking-film-heritage

2020 COVID-19 pandemic which led to the closure of many cultural institutions for many months and an acceleration of their digital offer.

Table 4-2 Culture and heritage values 2000-2019 using any non-market valuation methods, by asset type

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Archaeological asset	18	16	1	1	19	17
Archaeological asset	10	8	1	1	11	9
Ruin	8	8	0	0	8	8
Art engagement	0	0	4	3	5	3
Art engagement	0	0	4	3	5	3
Built heritage	9	7	15	6	24	13
Built heritage	4	4	7	6	11	10
Castle	2	2	7	0	9	2
Cinema	0	0	0	0	0	0
City	1	0	0	0	1	0
High street	0	0	0	0	0	0
Stately home	1	1	1	0	2	1
Town hall	0	0	0	0	0	0
Village	1	0	0	0	1	0
Cultural institution	55	35	33	19	87	54
Archive	7	1	3	0	9	1
Concert hall	3	0	0	0	3	0
Festival	5	4	2	2	7	6
Gallery	2	1	8	4	10	5
Library	10	10	4	2	14	12
Museum	27	18	16	11	43	29
Music venue	0	0	0	0	0	0
Theatre	1	1	0	0	1	1
Digital asset	2	1	6	4	8	5
Digital asset	0	0	0	0	0	0
Public service	2	1	6	4	8	5

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
broadcasting						
Industrial heritage	0	0	1	1	1	1
Aqueduct	0	0	0	0	0	0
Bridge	0	0	0	0	0	0
Canal	0	0	1	1	1	1
Dock	0	0	0	0	0	0
Factory	0	0	0	0	0	0
Harbour	0	0	0	0	0	0
Mine	0	0	0	0	0	0
Port	0	0	0	0	0	0
Quarry	0	0	0	0	0	0
Rail	0	0	0	0	0	0
Warehouse	0	0	0	0	0	0
Watermill	0	0	0	0	0	0
Waterway	0	0	0	0	0	0
Windmill	0	0	0	0	0	0
Historical amenities	13	10	7	7	20	17
Garden	2	2	2	2	4	4
Historical amenities	7	7	5	5	12	12
Monument	4	1	0	0	4	1
Plaque	0	0	0	0	0	0
Public art	0	0	0	0	0	0
Protected area	1	1	2	2	3	3
Protected area	1	1	0	0	1	1
Protected business	0	0	0	0	0	0
Protected city	0	0	2	2	2	2
Protected structure	0	0	0	0	0	0
Protected town	0	0	0	0	0	0
Protected transport	0	0	0	0	0	0
Religious asset	2	1	2	2	4	3
Religious asset	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Cathedral	2	1	2	2	4	3
Total	100	71	71	45	171	116

5.1.5 Country of origin

The REA includes 171 values from 22 OECD countries. The United Kingdom is by the far the most common country with 76 values. Australia, Spain, USA are the next most common with 16, 16 and 13 values each. Sweden and Greece then have 9 values, with the remainder of the countries having 5 values or less.

Table 4-3 Culture and heritage values 2000-2019 using any non-market valuation methods, by country of origin

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Australia	7	7	9	3	16	10
Austria	0	0	1	1	1	1
Canada	1	0	1	1	2	1
Czech Republic	3	3	0	0	3	3
Denmark	3	3	0	0	3	3
Finland	2	0	0	0	2	0
Germany	0	0	1	1	1	1
Greece	9	8	0	0	9	8
Ireland	1	1	1	1	2	2
Italy	4	4	1	1	5	5
Lithuania	3	3	0	0	3	3
Mexico	1	1	0	0	1	1
Netherlands	2	2	0	0	2	2
Norway	2	2	0	0	2	2
Poland	1	1	0	0	1	1
Romania	0	0	1	1	1	1
South Africa	2	2	0	0	2	2

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Spain	16	10	0	0	16	10
Sweden	8	0	1	1	9	1
Turkey	1	1	0	0	1	1
United Kingdom	21	15	55	35	76	50
United States	13	8	0	0	13	8
Total	100	71	71	45	171	116

In the sections below, the REA results are written up for each non-market methodology in terms of:

- Thematic split;
- Methodological considerations;
- And Quality Criteria Rating (established by methodology)

5.2 Stated Preference: Contingent Valuation

CV surveys capture direct and indirect (pride and social interaction) use values and non-use (existence, altruistic, bequest) values. This includes option values where respondents may hold a value for the site if they wish to visit in the future.

SP: CV: Conclusion Preview:

While CV was the most popular non-market valuation method for culture and heritage assets, only 9 of the literature values were rated as high-quality. Most CV methods employed high-quality standards, such as incentive-compatible elicitation methods (single- or double-bounded dichotomous-choice rather than open-ended WTP elicitation methods). Dependent on the characteristic of the good valued, payment vehicles were compulsory or voluntary. This is acceptable as some cultural heritage assets are non-excludable (i.e., a historic town centre) and are not amenable to a compulsory payment vehicle. Uncertainty around the payment term or whether the payment should be made at individual or household level present major design flaws, which may lead to inflated net present values when aggregated and incorporated into business cases. These were present even in the peer-reviewed literature. This suggests

that there is significant room for improvement, particularly in the review process, in CV practices.

5.2.1 SP: CV: Thematic split

The most common category of culture and heritage asset was **cultural institutions** (n=62 values, of which 36 rated medium-high quality). **The most common cultural institutions valued using CV were museums** (13 rated medium-high quality), libraries (11 rated medium-high quality), and festivals (6 rated medium-high quality) (full table of results in Appendix Table 6-2).

Built heritage (7 values rated medium-high quality), archaeological assets (6 values rated medium-high quality), and historical amenities (5 values rated medium-high quality) are the next most common categories for which CV values have been estimated. Within these categories there are some notable differences between the academic literature (which provides 5 of the 6 good quality values for archaeological assets, with a more equal split in values for built heritage and historical amenities between academic and grey literature.

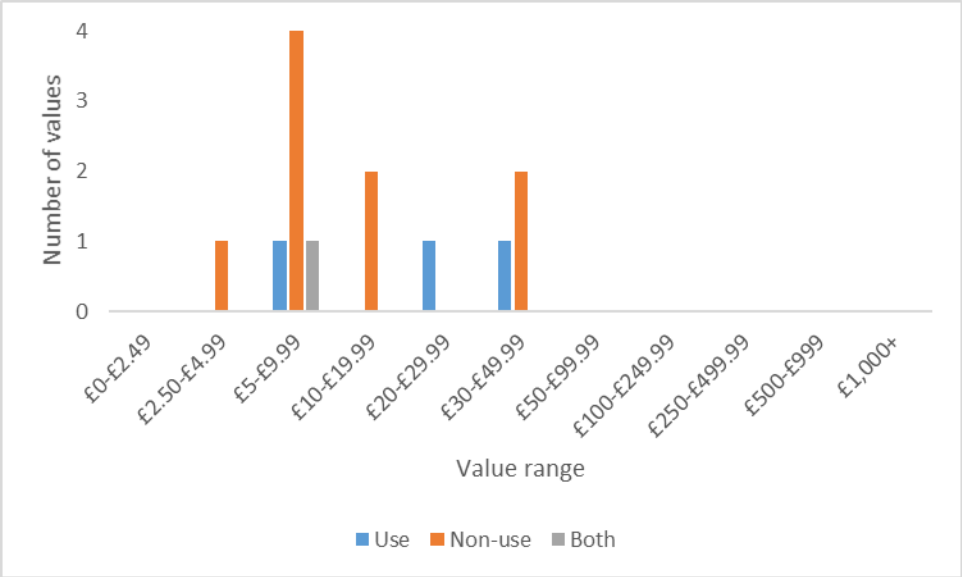
Three CV values for **digital assets** of medium-high quality were found in the literature (2 in grey literature and 1 in academic literature). In all cases, there were CV values of Public Service Broadcasting in the UK and other European countries. To pre-empt our results, the remaining values come from the other SP methodology of DCEs (n=2).

No CV values were identified for industrial heritage, and **only one for religious assets** (although that category is covered somewhat through BT (see Section 4.6). This suggests that there is an evidence gap for industrial heritage and religious assets as well, which needs to be filled using any of the available range of non-market valuation methods.

Cultural institutions that are accessed for cultural services, such as collections or performances, are the most common culture and heritage asset to be valued using CV methods.

The REA identified 22 CV values for **museums**, of which 13 were rated medium-high quality.

Figure 4-2 Willingness to Pay Range: Museum CV values (medium-high quality)



- **Museum WTP values were all within the range of £3.02 from visitors at the Natural History Museum to support its research and conservation work (a non-use value) and £42.87 for the conservation and maintenance of the Museo Patio Herreriano.**
- There were a smaller number of **use values for museums (n=3), and these ranged between £7.22** as a visitor entry fee for access to the Natural History Museum (on a sample of 616 visitors), to **£37.05** as an annual donation for preservation and maintenance of the National Museum of Sculpture, Spain (sample of 1,108 visitors).[18]
- **There were a higher number of non-visitor non-use museum values (n=8), ranging** from **£3.02** for research and conservation at Natural History Museum (surveying 616 visitors), to **£43.55** for the Museo Patio Herreriano de Arte Contemporaneo Espanol in Valladolid. We note that this high WTP should be seen as an outlier non-use value, as it was based on a sample of 287 culturally engaged non-users elicited by surveying visitors to a separate Madrid International Contemporary Arts Fair.[19] This sample would be expected to be subject to selection effects, whereby more culturally engaged individuals are more likely to 'select into' visiting a separate cultural festival, leading to an overinflated non-use value, demonstrating the danger of introducing overestimation through biased survey sampling strategies. Discounting the two values contributed for the Museo Patio Herreriano de Arte Contemporaneo Espanol in Valladolid,

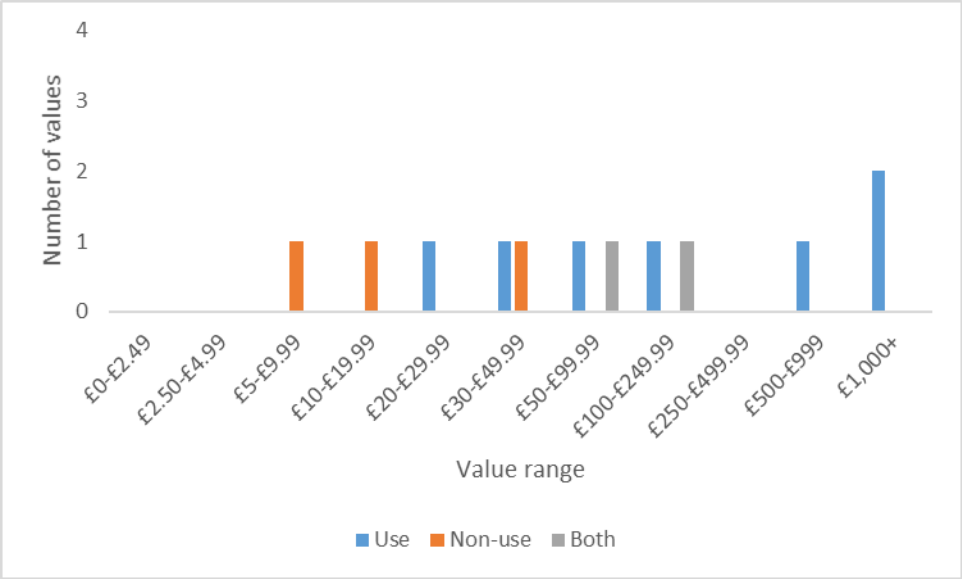
[18] Sanz et al. 2003
[19] Sanz et al. 2003

the range of non-use WTP values is between £2.50 and £19.99, which is more within the expected bounds, given that non-users in general would be expected to hold lower values than visitors for cultural institutions.

- **The REA also identified two museum WTP values ranging between** £7.48 and £7.74, consisting of a per visit WTA compensation value for closure of the Natural History Museum and Tate Liverpool respectively that encompassed both use and non-use elements.[20]

The REA identified 13 CV values for **libraries and archives**, of which 11 were rated medium-high quality.

Figure 4-3 Willingness to Pay Range: Library and archive CV values (medium-high quality)



- **Library/archive WTP values has a higher range than museums, of between £9.06 as a non-use value to maintain the British Library, for the general public and £4,712 WTA use value to sell a reader pass if reader passes were no longer issued, for users of the reading rooms.**
- **Library/archive user/visitor WTP/WTA values** range from £21.19 WTP to maintain current services at the library users closest library (self-identified within an online sample of 1,250 English library users)[21], to £4,712 as a minimum annual WTA to sell a reader pass if reader passes were no longer issued, among

[20] Bakhshi et al. 2015

[21] Daniel Fujiwara et al. 2019

229 users of the British Library's reading rooms. [22] The wide range in observed values here may partly reflect the disparity found between WTP and WTA, whereby WTA values are commonly found to be many magnitudes greater than WTP for the same good. This draws into question the reliability of the British Library user values obtained in that study.[23]

- **Library non-user/visitor WTP values** range from £9.06 as an annual tax to maintain the British Library (sample of 2,030 non-library users in the general public from all regions of the UK)) [24], to £48.97 as a one-off donation to support making the Contemporary Art archives and Collections of the Faculty of Fine Arts in Cuenca, Spain available to the public.[25]
- **The REA also identified two library/archive WTP values ranging between** £95.85 WTP to support maintaining all local public library services in Norway and prevent the libraries from closing down, and £242.10 WTA in compensation in response to the closure of the local library. In the former case, the high WTP value is likely driven by the fact that all libraries in the country are being valued by the general population sample. This is generally not a recommended approach in CV surveys as it introduces scope effects, where respondents are insensitive to a WTP to maintain 100 vs 1,000 libraries across the country, most of which they will never use. In the latter case, the high WTA may be driven by endowment effects and other cognitive biases that on average leads to a threefold disparity between WTA and WTP.[26] In both instances, the values are labelled as both use and non-use because they are elicited from the general public for libraries that they may or may not directly use themselves.

Other CV values for cultural institutions include:

- **Festivals**: 6 peer-reviewed WTP values, all rated medium-high quality. **Visitor values range from £8.70 as a tax** over and above any ticket prices paid for the Bradford Literature Festival (sample of 692 attendees)[27] to **£63.03 for a season ticket** to attend all of the concerts at the festival for locals at a classical music festival in the city of Santiago de Compostela (Spain), among local residents attending the festival.[28]

---

[22] Pung et al. 2004
[23] Pung et al. 2004
[24] Pung et al. 2004
[25] Saz-Salazar et al. 2017
[26] Y. Kim et al. 2015
[27] Goodspeed 2017
[28] Herrero et al. 2011

- **There was one festival non-visitor value of £4.98 as a tax** over and above any ticket prices paid for the Bradford Literature Festival (sample of 391 non-attendee local residents) [29] and **two WTP tax values** representing both use and non-use values for local residents (both attendees and non-attendees to prevent the festival from reducing in size by 25% for the National Arts Festival of Grahamstown USA (WTP of £5.63) the Klein Karoo Nationale Kunstefees festival in Oudtshoorn in South Africa (WTP of £7.33).[30]
- **Galleries:** 5 WTP values of medium-high quality, including use values for visitors to Tate Liverpool (WTP £11.76 as a one-off donation)[31] and the National Galleries Scotland (WTP £209.92 as an annual tax over an indefinite period for 222 visitors).[32] The wide range in the values obtained for two galleries in the UK likely demonstrates the insensitivity to payment term that can be introduced in CV surveys which fail to account for the fact that respondents may not distinguish between monthly and annual payments, and require an explicit payment term end date to make their stated preference realistic to their budget constraints.
- **Non-use values for galleries** included £21.18 among non-visitor individuals or businesses as a donor or event sponsor to support the Salamanca Arts Centre. .

In the heritage field, the REA identified **7 built heritage, 6 archaeological asset values, 5 historical amenity values, 1 protected area** and **1 religious asset value** medium-high quality CV studies. These findings suggest that CV is commonly applied to valuing heritage assets. The number of high-quality heritage studies suggests that the CV method is well-suited to historical assets and that the methodology has developed to be quite sophisticated when applied to this field. Indeed, some of the earliest CV studies in the culture and heritage sector were on historic townscapes and cathedrals, as identified in the Noonan review. Notable heritage valuations include:

- **7 built heritage WTP values** including 3 WTP values as an annual tax to preserve the built heritage interiors of stately homes in Europe from climate change damages in Italy (£44.61), Romania (£5.87) and Sweden (£4.34). The wide range of values for similar heritage assets across these three studies, all from the Mourato et al. 2014 study[33], shows the importance of BT techniques for avoiding the influence of outliers, and the Mourato et al. demonstrates this by surveying multiple sites in some countries

[29] Goodspeed 2017
[30] Snowball 2005
[31] Bakhshi et al. 2015
[32] Stevenson 2013
[33] Mourato et al. 2014

to asses transfer error between them. The REA observed **3 castle/stately Homes WTP values** from the same study surveying local residents, all report non-use values ranging from £5.27 as a one-off WTP for the conservation and preservation of the Raudondvaris manor, £22.54 for the renovation and adaptation of the Raudondvaris manor, and £40.40 as an annual donation for a programme of preservation of all manor residencies of the Kaunas region.[34]

- **6 archaeological asset/ruins WTP values**, from £7.32 per year over 2 years for changing the road layout within the Stonehenge World Heritage Site (the construction of a 2km tunnel for the A303 and the closing and dismantling of the A344) based on a combined visitor and general population sample[35], and £16.18 as a one-off donation from visitors to the archaeological site of Gobeklitepe in Turkey, to £74.23 as a one-off donation to rebuild an Arabic tower in Spain elicited as a non-use value from the local population.[36]

No academic-peer-reviewed industrial heritage values were found. Note that industrial heritage is not confined to post-eighteenth century structures but can include aqueducts and viaducts that date to Roman times. No values were found for this category of heritage in the REA, suggesting a research gap.

Finally, three CV values were identified in the REA for Public Service broadcasting, ranging £79.28 as an annualised UK household license fee to continue to access current BBC services[37], to £99.79 as an annualised license fee for Austrian public-service broadcasting on a sample of 722 students and graduates of the FHWien University of Applied Sciences, Vienna[38], to £258.68 as an annualised subscription service to the Irish public service broadcaster Raidió Teilifís Éireann (RTE) in replacement of the current licence fee.[39]

5.2.2 SP: CV: Methodology

It is standard for reviews in the valuation field to take stock of the 'state of the art' in terms of methodological approaches to eliciting value. In the case of CV surveys, these typically relate to elicitation methods, payment vehicles and terms, aggregation factors, and sample size. A high-quality study would address all these issues. These concerns are discussed in relation to methods employed within the literature in the following sections.

---

[34] Gražulevičiūtė-Vileniškė et al. 2011
[35] Maddison and Mourato 2001
[36] Del Saz Salazar and Montagud Marques 2005
[37] DCMS 2006
[38] Reiter et al. 2017
[39] Delaney and O'Toole 2004

Potential biases might overestimate the value of the site, if not adequately addressed in the methodology. For example, inaccurate preferences may be elicited in the hypothetical scenarios by survey respondents who alter their true preferences. These can include meta-preferences (the preferences people want themselves to have), rationalized preferences (more deliberated, internally consistent preferences), or laundered preferences (e.g., omitting "dirty" preferences).

5.2.2.1 SP: CV: Elicitation method

Elicitation methods range from open-ended questions to payment ladder or card approaches. Open-ended methods are the simplest direct elicitation technique where respondents simply state their willingness to pay. This can introduce large non-response rates, protest bids, and outliers that skew the data. As such, open-ended elicitation is not identified as good practice since at least the early 2000s[40].

There is some concern in the REA results that open-ended elicitation is still being used for eliciting CV values, with 27 such values for culture and heritage identified since 2000 (full table of results in Appendix Table 6-3). These values are mostly rated as low quality according to the REA Quality Grading Criteria, but in a very small number of instances (n=4), other survey design elements overcome these methodological limitations to results in a better overall grade. Nearly twice as many open-ended values were produced in grey literature (n=18) than the academic literature (n=10). This may suggest that the best practice message about the lack of robustness and high risk of bias introduced by open-ended methods has not been fully internalised by the research community in the culture and heritage sector. Our recommendation is that clear guidance be provided to analysts and culture and heritage institutions who may be commissioning research that open-ended elicitation of WTP should be avoided except in very rare cases.

Payment ladder or card approaches present respondents with a range of monetary amounts. This method eliminates anchoring (and starting point) bias and provides a visual aid to the cognitive process of valuing the good. However, the range of values presented may bias willingness to pay values given by the respondent. The inclusion of an open-ended 'other amount' option helps to reduce this effect.

The payment ladder was the most common single elicitation method (29, of which 23 medium-high quality values). This was broadly split between academic and grey literature (8 and 15 medium-high quality rated values respectively). This suggests that payment ladders remain popular for studies of all types in the culture and heritage sector, enabling the elicitation of WTP values in a

[40] Bateman et al. 2002

relatively easy to design way, while still producing values of good quality rating according to the Quality Grading Criteria used in this REA.

Higher quality single-bounded or double-bounded dichotomous choice questions allow the researcher to randomly present monetary figures to the respondent, with a follow-up figure presented in double-bounded dichotomous choice. Both methods are thought to ease any cognitive burden on respondents and provide incentives for reporting their true willingness to pay values. However, these higher quality approaches require larger sample sizes and may also encourage a larger number of protest bids than their open-ended counterparts.

In combination, dichotomous choice is the most commonly used method, with 32 CV values, either single-bounded (n=10) or double-bounded (n=20), or unspecified (n=2). Dichotomous choice was much more commonly used in the academic literature with 28 of the dichotomous choice values coming from peer reviewed academic papers, and only 4 coming from grey literature, of which 2 did not clearly specify whether they used single or double-bounded dichotomous approaches. This may suggest that the greater complexity in designing and analysing dichotomous choice WTP surveys and the higher and more costly sample size required form a barrier to entry that most grey literature reports are unable or unwilling to take on. As such, it tends to be only in academic literature that the high incentive-compatibility of dichotomous choice methods is considered necessary. Grey literature is more likely to make the trade-off between practical and theoretical considerations, meaning that in these cases a payment card may instead be chosen, which still fulfils the medium quality rating criteria for CV studies in this REA.

5.2.2.2 SP: CV: Payment vehicle

To elicit willingness to pay values, payment vehicles can either be compulsory or voluntary. Compulsory payment vehicles, such as taxes and entry fees, force consequentiality on the respondent to provide a hypothetical willingness to pay value. As such, these payment vehicles should be adopted wherever possible to increase the methodological robustness of SP studies. Conversely, voluntary payment vehicles, such as donations, have weaker consequentiality and allow for free riding, thereby possibly reducing the reliability of the results. Although compulsory payment vehicles are generally preferred as a means to reduce hypothetical bias, CV surveys also have to be responsive to the real-world, meaning that voluntary donations will often be the only payment vehicle choice available, and this should not be seen as preventing good quality CV design.

The majority of values (rated medium to high quality) applied a compulsory payment mechanism such as a tax (n=24), license fee (n=2) entry/membership fee (n=6), or other obligatory payments or subscriptions (n=3) (results in Appendix Table 6-4). The tax payment mechanism was most commonly applied in the academic literature, with nearly twice as many good quality WTP tax

values (n=15) compared to the grey literature (n=9). All license fee payment vehicles are found in the grey literature, which stems from the high number of values for public service broadcasting which are valued in grey literature reports.

However, almost as many good quality values were based on voluntary donations (n=18), and the majority of these are within the academic literature (n=11) compared to the grey literature (n=7). Therefore, while we can say that compulsory payment vehicles are preferred for eliciting WTP for culture and heritage assets, there are also a substantial number of voluntary donations. This likely reflects the different contexts in which culture and heritage assets are valued, which may preclude the use of compulsory payment vehicles, meaning that donations can be appropriate for many culture and heritage asset types. In many cases, in particular when studying cultural goods and services, compulsory payment mechanisms such as entry fees and taxes may not be appropriate, due to political sensitivities of institutions and funding bodies to a hypothetical discussion of raising such payment arrangements for cultural institutions. Indeed, in such cases, donations are the most common ways in which cultural institutions raise additional funding for the public goods they provide.[41]

One value in the academic literature did not specify the payment vehicle which is somewhat concerning as it suggests that the researcher were not conscious of the hypothetical bias that can be introduced by poorly defined and unrealistic surveys, giving much lower confidence in the quality of those values, but thankfully this study was very much in the minority. Only a minority (n=5) of studies applied a Willingness to Accept compensation payment vehicle. This may be a reflection of the challenges that have been found with WTA, notably its tendency towards overstated values (typically around 3 to 4 times that of WTP for the same goods)[42] and its encouragement of non-consequential responses, whereby the respondent thinks "if compensation is on offer, I would like some".

5.2.2.3 SP: CV: Payment term

Similarly to payment vehicles, payment terms, whether one-off or recurring, should be specified clearly in the survey. Choice of payment term should be informed by the policy context, but unlimited recurring periods should be avoided (unless accompanied by follow up questions asking how long respondents would actually pay for). This is because it assumes a continuous payment of that WTP year on year, which respondents may not actually be willing to do, but to which they were not explicitly informed. For instance, Kim and Haab[43] found that stating a willingness to pay value for a recurring period does not vary significantly across projects of varying lengths, which

[41] Lawton et al. 2019
[42] Brown and Gregory 1999; Horowitz and McConnell 2002; Y. Kim et al. 2015
[43] S.-I. Kim and Haab 2009

is strongly suggestive of an insensitivity to scope among CV survey respondents when it comes to payment term. To avoid this, one-off payments or annual payment over a fixed period will reduce the risk of payment term-inflation over a multi-year evaluation.

Fixed-term payment terms should always be adopted where possible, and in many cases in the literature on valuation of cultural and heritage assets between 2000 and 2019 this does occur. 28 good quality values applied wither a one-off payment (n=18) a fixed number of years (n=4) or as a per visit or per season WTP (n=6, all entry fee mechanisms) (Appendix Table 6-5).

However, a number of values (n=31) adopt an indefinite recurring payment or do not specify the payment term. This can lead to significant problems when analysts attempt to aggregate culture and heritage WTP values as a present value across an evaluation period of greater than one year, as is standard in economic business case analysis, because it assumes that all individuals will pay that amount year on year, which may not be the case and only be an artifact of the fact that the CV survey has been designed without a clear payment end date, thus failing to properly engage the respondent's thinking around their personal budget and implicit discounting of future benefits.

Of more concern, this tendency towards poor practice in not clearly defining a payment term appears to be more prevalent in academic studies (n=22 values) than the grey literature (n=9 values) even among studies that were rated as good quality according to the REA Quality Grading Criteria.

5.2.2.4 SP: CV: Aggregation factors: Individual or household

The imprecise definition of factors relevant to aggregation, such as payment term or payment at household level, can lead to inaccurate calculation of net present values over a 30-year evaluation period. Other factors that can lead to information loss and overstatement of values in the aggregation relate to whether values are elicited on behalf of the individual or household since the analyst needs to know whether values should be aggregated by the total adult population or the total number of households in a region. If this information is not specified then information loss occurs, and this can introduce inaccuracies in the aggregation process.

Valuation scenarios should elicit payment at household level and aggregate to number of households in relevant population. Even though some respondents may respond in follow up questions that they thought of the payment as an individual, it is recommended to err on the side of caution with a household aggregation, as this will reduce the risk of over-estimation. Equivalisation of WTP by number of people in the household is technically feasible but will likely lead to inflated WTP values.

A higher number of good quality values elicited WTP as an individual (n=29) rather than household (n=7) (Appendix Table 6-6). This may be partially driven by the fact that many of the culture and heritage values are elicited as a hypothetical WTP to continue to access sites which are currently free. Under these scenarios, WTP will be more closely related to individual experiencing the asset, rather than the shared value of the household.

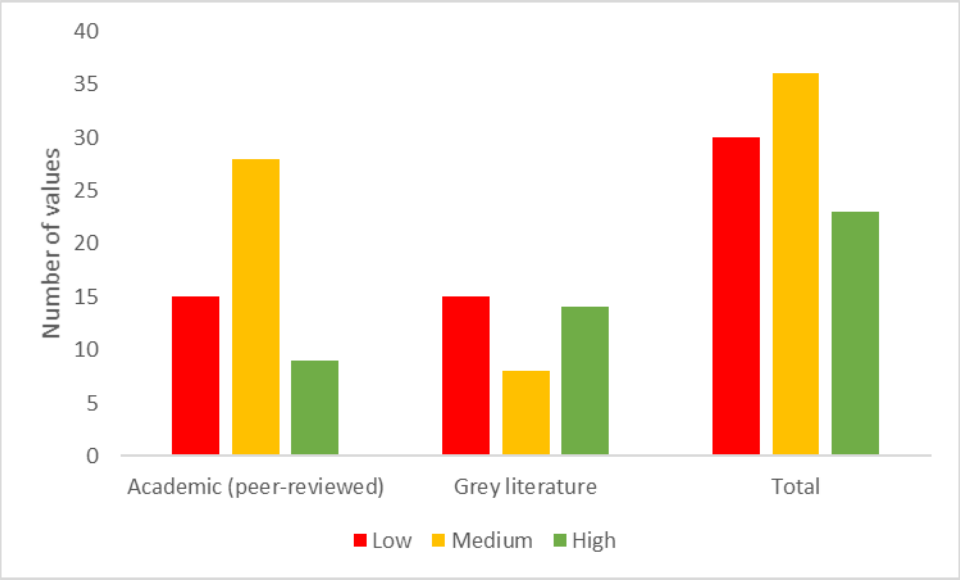
A surprisingly high number of values were elicited without clearly defining whether the payment was on behalf of the individual or household. 23 values were elicited in this way. Although the Quality Criteria Rating ranked these values as medium to high quality, on account of their other design characteristics, these values should nonetheless be taken with caution of introducing the risk of overstatement of aggregate value. This problem was prevalent even in the peer-reviewed academic literature.

5.2.3 SP: CV: Quality Criteria Rating (by methodology)

When comparing the source of papers applying CV to value culture and heritage assets between academic and grey literature, we see that out of a total of 52, the majority of academic CV values were rated medium quality (n=28), with 15 rated low quality and only 9 rated high quality. While these 9 papers were graded high quality, there is still room for improvement in the literature as some studies reported low samples, open-ended payment elicitation question, and lack of consequentiality checks.

In contrast, while a higher proportion of CV values in the grey literature were rated low quality (15 out of 37), of the remainder, a higher number (n=14) were rated high quality, with only 8 rated medium quality.

Figure 4-4 Quality Criteria Rating, Stated Preference: Contingent Valuation



5.2.4 SP: CV: Conclusion

The most popular method for non-market valuation of culture and heritage assets in the academic literature was CV. While a favourite for cultural site valuation, only 9 of these values were of high-quality, suggesting that there is significant room for improvement in CV practices.

In terms of methodological considerations, *the majority of studies are setting a high-quality standard in the design of CV surveys by using the most incentive-compatible elicitation method available that of dichotomous-choice whether it be single-bounded or double-bounded. A substantial number still apply the payment ladder. This may be because in situations where sample size is lower and there is an expectation of a large number of zero responses, dichotomous choice is more demanding, requiring a larger sample to present enough bid levels to capture the range of WTP, and possibly leading to imprecise value estimation if insignificant sample size is available. The payment card method has been shown to strike a balance between theoretical properties and such practical constraints. Note that both methods are approved in the best practice literature as creating an 'incentive compatible' payment scenario which most closely matches the contexts in which consumers make real decisions in actual markets.*

A concerning number of values applied open-ended elicitation methods, which are subject to considerable methodological limitations and biases. Clear guidance should be provided to analysts and cultural institutions who may be commissioning research that open-ended elicitation of WTP should be avoided except in very rare cases.

CV values are split between compulsory payment vehicles (such as entry fees and taxes which force consequentiality) and voluntary payment mechanisms (like donations). This may be due to inherent characteristics of goods and services in the cultural heritage sector. Certain types of cultural heritage assets are non-excludable, such as a historic town centre or greater landscape, and are not amenable to a compulsory payment vehicle. In such circumstances, voluntary donation payment vehicles are acceptable and still provide good quality values.

The REA identified a number of persistent design flaws in CV studies that can be problematic when WTP/WTA values are aggregated and incorporated into business cases. These include uncertainty around payment term or payment at household level, which can lead to inflated net present values. A surprisingly high number of values were elicited without clearly defining whether the payment was on behalf of the individual or household. Any attempt to aggregate them to total population or household numbers will introduce some information loss, introducing the risk of overstatement of aggregate value. This problem was prevalent even in the peer-reviewed academic literature. This is something that we think should be taken into account more in the peer review process, to ensure that researchers are thinking ahead to

SP: DC: Conclusion Preview:

A common valuation method in the literature with most values given medium to high-quality ratings, the DCE method presents a sophisticated way of eliciting WTP for various attributes of service delivery (e.g., expansion of services, improvements to congestion, and provision of ancillary services like restaurants and cafes). Added values, such as the preservation and discovery of heritage assets with potential value for future generations, can be elicited amongst visitors and the general public. Because of this, DCE is well-suited to eliciting both use and non-use attributes of cultural heritage sites. However, a DCE research gap was found, with few values elicited for built heritage and no values identified for art engagement, industrial heritage, protected areas, or religious assets. Most values in the literature employed a balanced approach between complexity and completeness in their design (i.e., less than 8 attributes with less than 4 levels in each). A sub-sample of grey literature studies were of low-quality due to poorly designed surveys and low sample sizes. While DCE methods present trade-offs in the provision of different attributes of a cultural service, a simpler CV method is likely more cost-effective and appropriate if the valuation is eliciting the value of a site as a whole. Caution should be taken in the application of values with an

potential pitfalls that their work could introduce when applied by practitioners, and designing their CV surveys to avoid such uncertainties.

5.3 Stated Preference: Discrete Choice

Similar to CV, Discrete Choice methods elicit direct and indirect use values and non-use values for a good. While CV presents respondents with a hypothetical scenario to state a value for, Discrete Choice methods present a series of alternatives (with varying attributes and levels) from which respondents select their preferred option. The result is a complex decision-making scenario that, if done well, is reflective of the market.

5.3.1 SP: DC: Thematic split

The majority of values applying DCEs to culture and heritage assets between 2000-2019 related to cultural institutions (n=10) (full table of results in Appendix Table 6-8). All 10 values were rated medium-high quality and were provided by academic peer-reviewed studies, of which **9 values were for museums**. This was followed by **11 values for archaeological assets and ruins** (10 rated medium-high quality), all from academic literature. There were **3 values for built heritage**, all medium-high quality rated and within the academic literature.

Four DCE values were identified for digital assets, specifically public service broadcasting, all within the grey literature, of which 2 were rated medium-high quality. This is a similar finding to the CV section, with public

service broadcasting organisations and agencies commissioning SP analysis within grey literature reports to demonstrate the surplus value that public service broadcasting provides to society. We explore the application of DCE to public service broadcasting services further below.

No DCE values were identified for industrial heritage or religious assets. This suggests that there is an evidence gap for DCE valuation of urban built heritage (excluding archaeological assets) which can be filled by commissioning research in this area in the future.

The REA identified 9 DCE values for **museums**, all rated medium-high quality, and all from peer-reviewed academic literature. DCE elicits WTP values for different aspects of the museum experience. All 9 DCE studies elicited use values from visitors.

- **WTP for new service aspects:** Tourists were willing to pay £0.75 as a per-visit payment to have a restaurant/bar at the Heraklion Archaeological Museum.[44] Visitors to the British Museum would be willing to pay £1.53 if it only opened on a Thursday or a Friday for a group of specialized or core cultural users.[45] This is an unusual value, given that it represents a reduction in opening hours, but the authors note that it likely represents a preference on behalf of members to exclude some of the general public at designated times during the week. A per visit payment of £1.34 was offered for multimedia services at the Galleria Borghese museum, Rome.[46]
- **WTP for expansion of collections:** Visitors to the British Museum reported a per-visit WTP £2.09 for a 66% increase in the number of temporary exhibitions at museum.[47] Counterintuitively, WTP values were higher (at £2.40) for a smaller increase of 33%. Broadly similar results were found for the Galleria Borghese museum, Rome, where visitors were willing to pay an extra £3.42 on top of current entry fees for an unspecified increase in additional temporary exhibitions at the museum. [48] The Galleria Borghese shows the power of DCE in calculating the additional value that visitors would be willing to pay for hypothetical changes to the services offered at cultural institutions, which can even be calculated on top of existing entry fees charged there.
- **Non-use attributes:** Visitors to the Galleria Borghese museum, Rome were willing to pay £5.02 per visit for an increase in

[44] Alexandros and Jaffry 2005
[45] Jaffry and Apostolakis 2011
[46] Mazzanti 2003
[47] Jaffry and Apostolakis 2011
[48] Mazzanti 2003

conservation/restoration services.[49] This can be defined as a non-use value, since the beneficiaries are not the visitors, but for the continued existence of the cultural assets for current and future generations.

In the heritage field, the REA of peer-reviewed academic DCE studies identified **10 archaeological asset values** (6 ruins and 4 archaeological sites), and **3 built heritage values**, all of medium-high quality. These findings suggest that DCE is commonly applied to valuing heritage assets of different types.

10 values were for different service attributes of **archaeological asset/ruins**:

- **WTP for new service aspects:** Visitors to the Roman archaeological site of Vindolanda at Hadrian's Wall had a per visit WTP of £3.47 to see an increased amount of reconstructions (i.e., full-size replicas of Roman buildings) on the site.[50] Tourists were willing to pay £5.88 as a per-visit payment to have a restaurant/bar at the Knossos Palace in Crete[51], which is higher than the WTP value for a restaurant/bar at the Heraklion Archaeological Museum also valued in the same paper. *This may suggest that leisure amenities are valued more highly at heritage sites than museums, possibly due to the visual amenity of the site.*
- **WTP for expansion of collections:** Visitors to Vindolanda at Hadrian's Wall had a per-visit WTP of £10.22 for an increase (33%) of artefacts discovered at the site to be exhibited within the site's museum.[52] Lower results were found for the Galleria Borghese museum, Rome, where visitors were willing to pay £3.42 for an unspecified increase in additional temporary exhibitions at the museum.[53]
- **WTP for improvement in congestion:** Tourist visitors to Knossos Palace in Crete reported a per-visit WTP of £1.45 for a 50% improvement in congestion.[54]
- **WTP for different opening hours:** Tourist visitors to Knossos Palace in Crete reported a negative per-visit WTP of £1.45 for a 50% reduction in entry fees after 4pm and on Sundays.[55]

[49] Mazzanti 2003
[50] Kenneth G. Willis 2009
[51] Alexandros and Jaffry 2005
[52] Kenneth G. Willis 2009
[53] Mazzanti 2003
[54] Alexandros and Jaffry 2005
[55] Alexandros and Jaffry 2005

- **Non-use attributes:** Visitors to Vindolanda at Hadrian's Wall were willing to pay £14.60 per visit for excavations and research to continue at the site. [56]. Danish citizens were willing to pay more as an annual tax increase to ensure permanent preservation of archaeological artefacts from Stone Age villages buried within the topsoil through wetland restoration (£106.88) compared to reducing the destruction of wetlands, where only an additional £72.62 in annual income tax was offered.[57] The Danish general public was asked their WTP for an extension of the restoration area for archaeological artefacts from Stone Age villages currently buried within the topsoil (+£0.11 per 100 hectares). Overall WTP was £8.96 as an increase in annual household income tax for ensuring permanent protection of the Stone Age artefacts, or £12.17 WTP for reducing destruction of the site.[58]

Other culture and heritage DCE values included:

- **3 built heritage WTP values:** The Australian public were willing to pay more in increased annual household tax (no fixed end date) to extend the duration of temporary exhibitions by one month (+£0.94), for the hosting of various events (+£1.20), and to have conference rooms, a shop and café, and fine dining (also +£1.21). at the Old Parliament House, Australia.[59]
- **2 public service broadcasting WTP values:** For different combinations of television broadcasting in the UK. License fee holders were willing to pay £21.36 as an annualised household license fee for ITV1, Channel 4 and Five, and £96.80 for all public-service broadcasting including the BBC at the current licence fee in addition to the public service programming on ITV1, Channel 4 and Five).[60]

No DCE values were identified in the DCE for art engagement, industrial heritage, protected areas, or religious assets, suggesting that this is a research gap for which DCE could assist in understanding the value of different service aspects that are currently offered or could be improved from these culture and heritage assets.

---

[56] Kenneth G. Willis 2009. This can be broadly defined as a non-use value, since the beneficiaries are not the visitors, but the discovery and existence of the cultural assets now and for future generations. Although, there may be some knock-on option use value if the artefacts were to be displayed in the future

[57] Lundhede et al. 2013

[58] Lundhede et al. 2013

[59] Choi et al. 2010

[60] Ofcom 2008

5.3.2 SP: DC: Methodology

While experimentally complex, Discrete Choice designs offer a robust value provided a thorough empirical design is followed. To achieve this, welfare consistent designs, regression modelling, and incorporating zonal or individual datasets need to be considered. These are discussed below.

As this method elicits preferences based on hypothetical choices, bias and errors are likely. Respondents may correct their responses (meta-, rationalised, or laundered preferences), take survey shortcuts, present inconsistent responses, and show signs of fatigue. Any stated willingness to pay values may thereby be sensitive to the Discrete Choice design and might unintentionally direct respondents to the simpler alternative (i.e., status quo).

As such, there is a trade-off between complexity and completeness; a simpler range of hypothetical choices may not reflect the true market, but a complex range of choices might burden the respondent. Statistical tools help to reduce the number of hypothetical scenarios to a digestible level; however, the reduced scenarios may not capture all the important attributes for decision making.

5.3.2.1 SP: DC: Discrete Choice design

In order to obtain true preferences from respondents, hypothetical choices in Discrete Choice designs must avoid deliberated preferences such as meta-, rationalised, or laundered preferences. Discrete Choice designs must mimic realistic market choices, where the market choices are somewhat complex and complete, but simplify the number of choice scenarios (or levels) by using statistical techniques. An accurate balance of this trade-off will reduce any cognitive burden on respondents and, provided the choice scenarios are realistic enough, will obtain true preferences. Focus group, piloting, or prior estimates should influence the Discrete Choice design to ensure a realistic and high-quality design.

The REA shows that design of DCE studies for culture and heritage assets followed robust methodological approaches. There were no instances of simple models in either the academic or grey literature. The majority of DCE values were obtained using a small number attributes of an asset, with few levels so as not to cognitively overburden the respondent and be more market realistic (29 values, of which 25 rated medium-high quality) (Appendix Table 6-9). All of these methodologically robust values were produced in the academic literature

Of the 4 DCE values identified in the grey literature, only 2 were rated medium to high quality in the Quality Criteria Rating. All of these DCE designs contained more than 8 attributes and a high number of levels. Although this is not on its own cause for a low quality rating, it does

introduce the risk that the respondents were cognitively overburdened by the DCE exercise, which could lead to inaccurate WTP values. We recommend that future commissioning of DCE valuation in the grey literature should be aware of this risk of overburdening survey respondents with an excessive number of attributes, and advise analysts to exercise caution in the application of such values.

5.3.2.2 SP: DC: Regression modelling

Advanced regression modelling (e.g., nested logit, mixed logit, latent class models) capture preferences for attributes. Simpler regression modelling techniques, such as conditional logit and binary logit, are less robust. More attribute levels can be controlled through more advanced regression modelling.

The majority of studies applied advanced regression modelling (e.g., nested logit, mixed logit, latent class models) which are able to capture preferences for attributes, and can manage a higher number of attributes entered over blocks within the regression model, by controlling for attribute levels (Appendix Table 6-10). A smaller number of values applied the simple binary logit model, and were excluded from medium-high quality by the Quality Criteria Rating.

5.3.2.3 SP: DC: Welfare consistent design

As discussed in section 4.3.2.1, the Discrete Choice design influences the quality of the results. An efficient design (S or D efficient) or that using software (i.e., NGENE or SAS), ensures a welfare consistent approach by automatically grouping choice sets by a computer algorithm that maximises the model fit. Thereby gaining preferences for each model level without relying on respondent sampling size. These designs have correlated effect estimates, whereas orthogonal designs do not share this feature. Rather orthogonal designs assume and evaluate factors independently; thereby not accounting for the potential relationships between attributes.

The majority of studies report efficient design, which helps to ensure welfare consistency and increases the robustness of the WTP values obtained (Appendix Table 6-11). A number of studies did not provide details on the design underlying their values. Even though these values may be considered robust on other factors, audiences should as a minimum expect that efficient design protocols be followed as standard, given the easy availability of orthogonal and optimal design functions within common statistical software.

We recommend that reviewers of academic literature and those commissioning DCE research in the grey literature make reporting of efficiency design elements a requirement of publishing of DCE values going forward.

5.3.2.4 SP: DC: Welfare consistency

There are a number of Discrete Choice designs which are not considered full DCE because they are not consistent with how people form welfare preferences in real markets.

Bateman et al.[61], in their 2002 Manual of Economic Valuation with Stated Preference techniques assume that DCE should always have a status quo to ensure welfare consistency (see table below, from Bateman et al.). As described by Figure 4-5, Paired comparisons and contingent ranking or rating, for instance, do not cater for market realism as previously discussed in section 4.3.2.1. Furthermore, they do not allow for an opt-out option. However, other design elements are also important in ensuring that the DCE scenario is realistic.[62]

Table 10.1 Main choice modelling alternatives

Approach	Tasks	Estimates consistent with welfare economics?
Choice Experiments	Choose between (usually) two alternatives, versus the status quo	Yes
Contingent Ranking	Rank a series of alternatives	Depends[1]
Contingent Rating	Score alternative scenarios on a scale of 1–10	Doubtful
Paired Comparisons	Score pairs of scenarios on similar scale	Doubtful

1 In order to interpret the results in standard welfare economic terms, one of the options must always be currently feasible.

Figure 4-5 Main choice modelling alternatives, from Table 10.1 in Bateman et al. 2002

The inclusion of a status quo option provides respondents with a scenario that is consistent with the current scenario if these alternative options do not occur. It therefore allows respondents to select the current status quo to continue if they do not prefer the alternatives. Opt-out choices (e.g., "Don't know") allow respondents to opt out of choosing between attributes. These options are important to include as some visitors may not have a preference between the choices or may not visit the site in question and prefer not to provide a response.

The majority of academic DCE studies adopt welfare consistent design, which means that the values obtained can be considered welfare-consistent WTP values for the purposes of CBA (Appendix Table 6-12). In combination, this

[61] Bateman et al. 2002
[62] Greene 2007

suggests that DCE research in the cultural heritage sector is well-established and well-designed in the majority of cases, and capable of providing robust WTP values for specific attributes of cultural heritage assets for use in policy business cases.

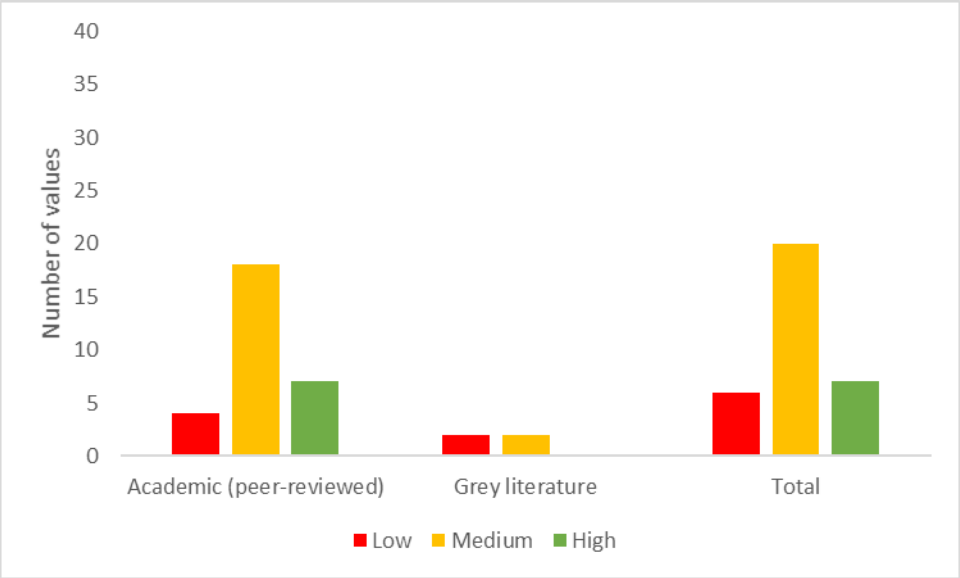
5.3.3 SP: DC: Quality Criteria Rating

Overall, the majority of DCE values (20 of 33) were rated as medium quality by the Quality Criteria Rating.

A small number of values (7) achieved high-quality ratings due to their structured Discrete Choice design, robust regression modelling, and market realistic scenarios. High quality values were all obtained from the academic literature.

6 DCE values were rated as low quality by the Quality Criteria Rating due to small sample sizes and poor designs (i.e no focus groups or pilot testing). A higher proportion of lower quality values came from the grey literature (2 of 4) than the academic literature (4 of 29). We suggest some reasons for this finding in the conclusion box below.

Figure 4-6 Quality Criteria Rating, Stated Preference: Discrete Choice



5.3.4 SP: DC: Conclusion

Similar to CV, Discrete Choice methods prove to be a commonly employed valuation technique in the literature, for both cultural and heritage sites. While most values were given medium to high-quality ratings, something that is concerning is that a sub-sample of these were of low-quality. Compared to CV, DCE methods require a larger sample size.

DCE is frequently used to elicit WTP for various attributes of service delivery, such as expansion of services, improvements to congestion, and provision of ancillary services like restaurants and cafes. As such, DCE is well-suited to eliciting WTP for both use and non-use attributes of heritage sites, in terms of both preservation and discovery of heritage assets with potential value for future generations, among both visitors and the general public.

The large number of values relating to different aspects of service provision at cultural institutions shows the power of DCE in calculating the additional value that visitors would be willing to pay for hypothetical changes to the services offered at cultural institutions, which can even be calculated on top of existing entry fees charged there.

DCE is used less commonly for built heritage (with no values for stately homes, in contrast to CV), or industrial heritage. There is no structural reason why DCE could not be applied in the same way to attributes of these cultural heritage assets among visitors if policy demand dictated.

No DCE values were identified in the DCE for art engagement, industrial heritage, protected areas, or religious assets, suggesting that this is a research gap for which DCE could assist in understanding the value of different service aspects that are currently offered or could be improved from these culture and heritage assets.

However, there is a general issue about whether a DCE is needed for all culture/heritage asset types. DCE is a more sophisticated approach than CV that is required for testing trade-offs between attributes. If the analyst is interested in varying the provision of different attributes of a cultural service, then DCE is worth considering. However, if the analyst is interested in the value of the asset as a whole, then a simpler CV is likely to be a more cost-effective and appropriate method.

In terms of methodological considerations, the vast majority of DCE papers adopted a balanced approach between complexity and completeness, reducing the burden on respondents by valuing a manageable number of attributes (under 8) with a small number of levels in each (under 4). The majority of these studies focused on providing WTP values for specific attributes of cultural heritage assets, suggesting that researchers are correctly identifying the research questions for which DCE is most appropriate. A higher proportion of grey literature values were based on poorly designed DCE surveys containing more than 8 attributes and a high number of levels, which risks overburdening the respondent and producing unreliable valuations. We recommend that future commissioning of DCE valuation in the grey literature should be aware of this risk of overburdening survey respondents with an excessive number of attributes, and advise practitioners and analysts to exercise caution in the application of such values.

The high number of DCE studies that are based on a sample size below the minimum recommended threshold of 500 suggests that sampling issues, whether cost or sample access, might prevent researchers from obtaining the minimum sample size targets set by the Quality Criteria Rating. However, we note that the minimum sample sizes recommended in the Quality Criteria Rating were intended as rules of thumb. Sample size need to be calculated on a case by case basis, factoring in target sample groups and maximum sample frame considerations. We would recommend a deeper review of the low-quality rated studies to ascertain whether appropriate sample size calculations were undertaken, and in such cases the quality rating could be increased to medium or high. However, this level of review is beyond the scope of the REA.

5.4 Revealed Preference: Travel Cost

Unlike SP methods (e.g., CV and DCE), TCM is unable to measure indirect use value, option value, and non-use value. This method focuses observable data on recreational visits to sites to understand direct user benefits estimated as per visit values, which can be totalled up by the number of visitors per year to establish an annual direct use value of the site. [63] As this method relies on recreational use values of a site visit, it technically elicits value of a trip to the site, rather than the value of the site itself.

RP: TC: Conclusion Preview:

The TCM method was applied to museums, gardens, and historical amenities in the literature but very few values were rated as medium or high quality. This high number of low-quality values could be explained by the many ongoing and unresolved criticisms in the valuation of cultural heritage assets. For example, TCM elicits the value of the trip to the site, rather than the cultural value produced by the site itself. As a result, it is difficult to disaggregate the value for a site from multi-purpose trips. The method relies on observable visitor numbers to sites and is data intensive (e.g., it requires the individuals' choice of sites, place of residence, socio-economic and demographic characteristics, frequency of visits to the site and similar sites, and trip cost information). Furthermore, this method cannot capture spillover benefits and non-use values from the general population who do not visit, so it can only elicit recreational use values. Overall, these low-quality values suggest that the

[63] Boardman et al. 2010

5.4.1 RP: TC: Thematic split

While far less common than SP or other RP methods in the literature, TCM is mostly used to value cultural institutions. The REA identified **9 values for cultural institutions**, but only 2 rated good (medium-high) quality. **Museums were the most common cultural institution** using TCM (7 values, of which 2 rated medium-high quality). Only **2 values were found for historical amenities**, both of medium-high quality, and **2 values for built heritage**, both of low quality. **No TCM values were found for industrial heritage which represents a clear gap in the literature** (full table of results in Appendix Table 6-14). In the UK, for instance, it may be possible to combine data on Travel Cost with GIS data on the presence of built and industrial heritage collected by Historic England, to fill this gap in the valuation literature by commissioning research in this area in the future.

The majority of TCM values were reported in the academic literature. Of those produced by the grey literature, all were related to cultural institutions, and specifically museums (2 values, both rated low quality).

Museums: The REA found 2 academic peer-reviewed values of medium-high quality applying TCM to museums. Values ranged from £36.29 consumer surplus per visit, from a sample for 1,067 visitors to National Museum and Research Centre of Altamira, Spain (attached to the UNESCO WHS Cave of Altamira)[64] to £70.33 as the average value of an annual seasonal ticket to any of the 108 Dutch museums from a sample of nearly 70,000 National Museum Card holders (a card that allows holders free access to 442 Dutch museums).[65]

Historical amenities: There was one TCM value of medium-high quality with an average WTP value of £16.24 (individual per-visit) for a battlefield trip for any of three historic battlefield sites maintained by the US National Park Service on a sample of 277 visitors.[66]

Historical gardens: An average value of £22.12 (per-trip) consumer surplus was elicited from single-site visitors to one of three Australian botanic gardens spread across three geographical states of Canberra, Melbourne, and Sydney on 1,139 visitors.[67]

Archaeological assets: There was one TCM value for visitors to the Poseidon Temple in Greece, but this was calculated as an annual aggregate amount of £45 million. This value is difficult to compare to the individual-level values obtained in other TCM studies. However, the annual value of £45 million for a single site seems very high, and may be driven by the low-

[64] Corruchaga and Monforte 2006
[65] Rouwendal and Boter 2009
[66] Melstrom 2014
[67] Mwebaze and Bennett 2012

quality zonal modelling method applied and small sample size of 150 (questionnaire respondents). The study also does account for multi-purpose trips.[68]

5.4.2 RP: TC: Methodology

In order to be graded high-quality, the primary issues in TCM designs to be accounted for are single versus multi-purpose travel, accounting for model covariates, zonal or individual datasets, and appropriate sample size. These issues are discussed below in relation to the academic and grey literature.

TCM presents numerous issues. First, respondents may value and enjoy the travel to the site itself. For example, a scenic train journey to a heritage site. As travel might be valued as a positive activity, there may be no costs incurred, so to speak. Second, any results obtained are sensitive to assumptions made about travel time. Often a fraction of the wage rate is used, which may not be validated in empirical studies (Atkinson et al. 2018). Third, this method requires a multitude of information such as the individuals' choice of site, place of residence, socio-economic and demographic characteristics, frequency of visits to the site and similar sites, and trip cost. This might be inaccurately recalled or forgotten by respondents. Fourth, travel to the site may be undertaken for more than one purpose (Parsons 2017). For example, a visit to a historical library may be untaken for the historical value the library presents and for the goods and services at the site. It is therefore difficult to disaggregate the value for the site in question from these multi-purpose trips.

5.4.2.1 RP: TC: Single/Multi-purpose travel

In TCM, while multiple purpose trips are difficult to disaggregate the value of the site in question, single-purpose site visits present disaggregation problems as it often ignores the other potential purposes (and values gained) from the greater location.

The Quality Criteria Rating assigns medium quality to studies which attempt to deal with this issue of multi-trips in any way but do not give further detail. A high quality rating is given to studies which deal with the issues of multi-trips using one or more of the techniques below, but we not that there is no widely accepted method for dealing with this limitation:

- Using only a part of the total travel cost, which corresponds to the additional expenditures made to visit the area in question from the last stopover;
- Distributing the total travel cost on the basis of the time spent by the visitor at each of the sites visited;

[68] Tourkolias et al. 2015

- Distributing the total travel cost on the basis of the welfare derived by each of the sites;
- Identifying the various profiles of multipurpose trips undertaken by the visitors of the site in question and estimating the demand curve and the recreational value for each of them, while the recreation value of the area in question is estimated at a later stage as a percentage of the recreational value of each type of the multipurpose trips identified.

In the academic literature, most values were calculated from single-purpose travel (11) with only 2 values provided from multi-purpose travel (Appendix Table 6-15). Of which only 3 of these values were graded as high-quality. Lower graded values made large assumptions in their single-purpose travel. For example, one value claimed that as the museum is a World Heritage site, it can be assumed that any visitors would be visited the site as the primary trip purpose. Another value attempted to control for multi-purpose visits by directly asking respondents what their main travel purpose was and dropping those who did not mention the valuation site.

In the grey literature, both of the 2 TCM values applied a simple single purpose travel model with no attempts to correct for multi-purpose trips and were rated low quality.

5.4.2.1 RP: TC: Model covariates

When running the regression models for TCM, the inclusion of covariates will reduce any risk of bias in value estimates. For example, site characteristics such as the visitor group size, hours spent at site and the age of visitors, should be included as covariates in any regression model.

Just under half (6 of 13) of the TCM values from the academic literature had no covariates or overly simple covariates in the function (Appendix Table 6-16). All of these values were graded low-quality. Failure to account for the underlying factors driving travel introduces unobserved variable bias, which can lead to overestimates of non-market values for cultural heritage assets. These values were subsequently assigned a low quality rating in the REA evidence bank. For policy purposes, analysts should only consider as robust TCM values that apply covariates to control for standard drivers of travel behaviour. In the grey literature, all two TCM values were insufficient in their modelling, containing no covariates.

5.4.2.2 RP: TC: Zonal/Individual data

Zonal TCM, which predominantly uses secondary data sources about people's travel behaviour, can be conducted quickly and at a low cost. Zonal TCM uses secondary datasets by estimating the proportion of visitors to the site from zones, whereas primary data relies on surveying visitors or logging card holder visits.

Primary data is the preferred dataset for TCM, although it relies on respondent information which may be inaccurately recalled (e.g., total costs incurred by travel, distance travelled). It is therefore recommended to supplement primary data with latent class models, instrumental variables, quasi-experimental designs, and integration with SP or Geographic Information System (GIS) data for a more rigorous dataset.

Nearly double the number of TCM values for cultural heritage assets applied zonal travel models rather than individual-level data. This is surprising, given that best practice guidance in non-market valuation advises in favour of more advanced choice models. These values were subsequently assigned a low quality rating in the REA evidence bank. For policy purposes, analysts should only consider as robust TCM values that use individual-level data on travel behaviour, or data collected through primary studies.

5.4.3 RP: TC: Quality Criteria Rating

When comparing the source of papers applying TCM to value cultural heritage assets between academic (peer-reviewed) and grey literature (which are not necessarily peer-reviewed), we see that the majority of TCM values (8 academic literature, 2 grey literature) were rated low quality, with only 3 rated medium quality (3 academic) and 2 rated high quality (both in the academic literature).

In part this is due to sample size limitations. Around a third of academic peer-reviewed TCM studies are based on a sample size below the minimum recommended threshold of 250, while the same number are only at the medium quality threshold for sample size. This may reflect the fact that the majority of studies rely on low quality zonal data, which may be subject to sample size restrictions.

Around a third of academic peer-reviewed TCM studies are based on a sample size below the minimum recommended threshold of 250, while the same number are only at the medium quality threshold for sample size. This may reflect the fact that the majority of studies rely on low quality zonal data, which may be subject to sample size restrictions (Appendix

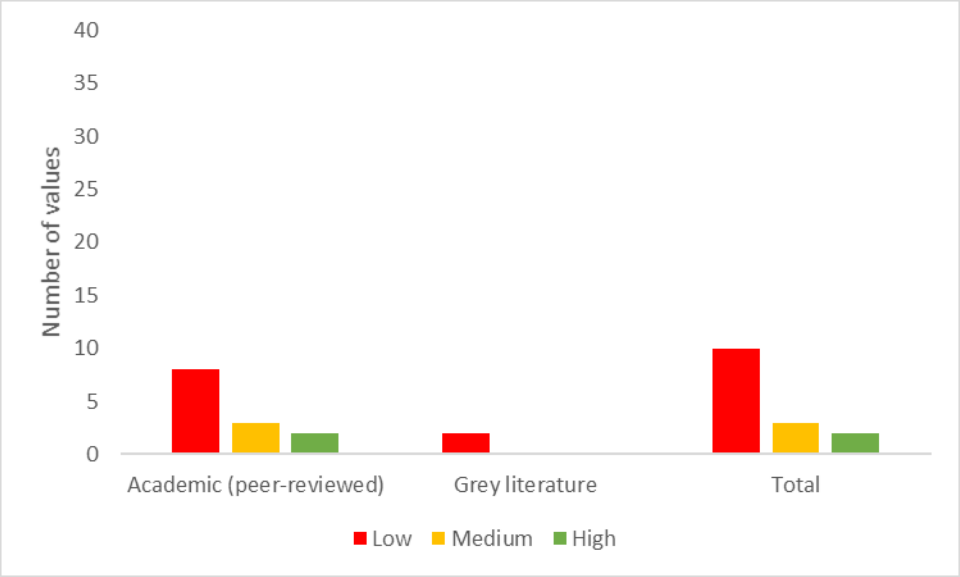
In the grey literature, one study did not specify its sample size, which is a major problem for analysts looking to assess the reliability and quality of culture and heritage values obtained through TCM.

Nine values in the academic literature obtained a minimum sample of 250 or more, with only 4 of these values of medium-high quality. One notable value was obtained using automated logging of individual visitor data, allowing for a very large sample size (69,643), which was accompanied with GIS data to present a robust model. Values with adequate sampling but poor methodological quality, and the four values which obtained an inadequate sample size, were

marked down. These included the employment of zonal data rather than individual data, or the lack of consideration into multi-purpose visits.

In the grey literature one TCM study was based on a large sample of 1,000+ and one study did not specify its sample size, which is a major problem for analysts looking to assess the reliability and quality of culture and heritage values obtained through TCM.

Figure 4-7 Quality Criteria Rating, Revealed Preference: Travel Cost Method



5.4.4 RP: TC: Conclusion

TCM does not present itself as a strong contender in the valuation of cultural and heritage sites, with limited high-quality values found in the literature. The disproportionate number of low-quality values may relate to the ongoing and unresolved criticisms of the TCM method. As outlined by Atkinson et al.[69], TCM is restricted to recreational use values related to visiting a cultural site, meaning that technically it elicits the 'value of a trip to the cultural site' rather than the 'cultural value produced by the site'. The REA therefore suggests that TCM values have a limited application which is reliant on observable visitor numbers to recreation sites. Consequently, the REA shows that TCM methods have been applied only to tangible culture and heritage assets like museums, gardens, or historical amenities being with observable visitor numbers.

Furthermore, this application is data intensive and requires data on individuals' choice of site, place of residence, socio-economic and demographic characteristics, frequency of visits to the site of interest and

[69] G Atkinson et al. 2018

other similar sites, as well as trip cost information, which can be difficult to collect in combination, as survey respondents may not be able to recall all of the data required for modelling, such as the travel costs incurred and distances travelled. It is also difficult to disaggregate the value of multiple purpose trips, where travel to the location of a culture and heritage site is undertaken for more than one purpose.[70]

The TCM method is also subject to significant methodological limitations, meaning that only a minority of the TCM values identified in the REA were rated as medium or high quality according to the rating criteria. TCM methods are unable to elicit non-use values from the general population who do not visit, and cannot capture the spillover benefits of cultural heritage assets to the local place.

5.5 Revealed Preference: Hedonic Pricing

The most commonly employed RP method for uncovering estimates of the value of nonmarket goods by using evidence of how people behave in the face of real choices is HP which involves examining people's purchasing decisions in markets related to the nonmarket good. It has commonly been applied using data from housing and labour markets. The intuition is that the price differential between otherwise identical houses that differ in their exposure levels to nonmarket goods and bads such as good schools, pollution, and crime, reveals information regarding people's WTP (WTA) for such goods.

In the culture and heritage sector, HP methods obtain direct use and indirect values (e.g., local benefits such as reduced crime rates), with option values

RP: HP: Conclusion Preview:

Hedonic Pricing is not well suited to the valuation of most cultural heritage assets, wherein benefits are often provided to visitors but not necessarily the price of houses nearby. Most HP values were obtained through primary (i.e., administrative data) with all marked as high-quality due to the large size and scope of the datasets. While not often used in cultural heritage valuation, there is an opportunity in valuation research to exploit enriched real estate data for assets, such as heritage buildings and protected urban areas where there may be numerous historical assets within the one location, that would affect house prices in the surrounding area. More often used in the grey literature, this method may be less popular due to the financial cost of using proprietary real estate data. While some administrative and real estate datasets are publicly available online, open-source real estate datasets would help to increase the use of this unique data method. This method provides robust valuations of

[70] Parsons 2017

for those who live nearby and wish to visit the site, but this method does not pick up non-use values. Values are established through house prices, where it is assumed that the closer to a site of interest the house is, the higher the prices. This method relies on the assumption that the proxy market is reflective of the real market and is therefore an accurate representation of preferences held for non-market goods.

5.5.1 RP: HP: Thematic split

HP was not a commonly used method in the literature, with only 5 values, of which 4 came from peer-reviewed academic studies. All hedonic values in the REA were rated as good (medium-high) quality. **HP was used to value historical assets**, such as **built heritage (n=1)**, **industrial heritage (n=1)**, and **historical amenities (n=3)** (full results in Appendix Table 6-19).

Historical amenities: A US hedonic study found that historic designation for an individual property was associated with an average property value premium of £4,570. However, another US study found that being adjacent to a historical area is not a positive benefit on average and is associated with an average lower house price of -£11,368. The same study found that for each 200-feet in distance from historical amenities, house prices increase by £38.40. However, we must account for the fact that there may be a number of unobservable factors which are endogenous to the historical character of the area that may not have been picked up in the model.[71] *Note, all values in the Hedonic section are treated as household level values, and are divided by the average household size of the country in which the study was based, to equivalise all values in the REA to individual-level values.*

Built and industrial heritage: House prices in the Dutch urban area of Zaanstad were found to be associated with listed heritage status of a building or historic cultural site in that area. Buyers are willing to pay an additional 26.9% for a listed building, with surrounding houses worth an extra 0.28 % for each additional listed building within a 50-m radius. Houses sold within a conservation area appear to gain a premium of 26.4% which confirms the existence of a 'historic ensemble' effect. The effect of listed heritage status for an average house in the municipality of Zaanstad was calculated at £25,879.[72]

[71] R. L. Hicks and Queen 2016
[72] Lazrak et al. 2014

Historic industrial assets: A 2019 study by a UK team lead by Peng et al. found that living within 100 metres of the canal and waterway network in Britain was associated with a small house price premium of £168.37.[73]

5.5.2 RP: HP: Methodology

A few considerations should be taken into account when practising advanced empirical design in HP models. These include accounting for model covariates, appropriate regression models, and comprehensive data sources. Each are discussed in turn below.

There is criticism that HP does not reflect the real housing market.[74] For one, transactions costs are high with limited supply in the housing market. HP also assumes free mobility. That is, individuals can adjust different levels of each characteristic of interest by simply moving properties without transaction costs. Further, property values reflect capitalised, rather than annual changes in value. Factors that drive house prices which are correlated with proximity of the house to the site of interest, need to be incorporated into the regression model for accurate valuation. If these concerns are not accounted for and accurate information of the market is assumed, omitted variable bias is likely to occur.

5.5.2.1 RP: HP: Model covariates

Within the regression models, covariates need to be included so as not to bias any value estimates. In HP Methods, covariates may include structural (e.g., house size and age) and spatial (e.g., neighbourhood attributes) housing characteristics.

The REA demonstrates that the majority of hedonic values used a good combination of structural and spatial covariates in their regression models (Appendix Table 6-20), giving greater confidence in the robustness of this method as applied to the cultural heritage sector.

5.5.2.2 RP: HP: Regression model

Advanced modelling techniques (e.g., instrumental variables, spatial or temporal fixed effects) are used to robustly estimate implicit price changes through effects on house prices or alternatively, full demand function for individuals in housing market. Regression models should achieve experimental causality through randomisation, control group comparison, and temporal effects.

The majority of hedonic values obtained a high-quality grading in their modelling, using regression models that, for instance, accounted for

[73] Peng et al. 2019

[74] Chau and Chin 2003

heterogeneity of the error structure with log-linear and semi-log functions for robustness (Appendix Table 6-21). This gives greater confidence in the robustness of this method as applied to the cultural heritage sector.

In the REA all values applied an OLS regression, with 3 values based on log linear OLS to account for heterogeneity of the error structure, and one value from the grey literature based on OLS plus Difference in Difference.

5.5.2.3 RP: HP: Data source

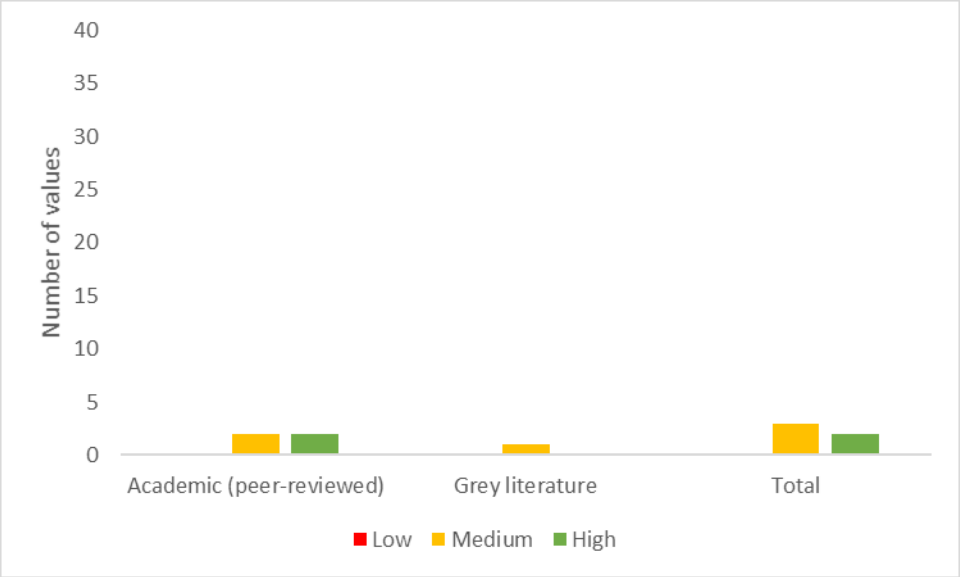
Values obtained using primary datasets or large datasets (e.g., real-estate transactions and listed prices) are generally considered to be less comprehensive than large datasets used in combination with alternative data (e.g., SP or GIS data). These combined datasets provide a richer data source for the value estimate. However, the data are often proprietary, and can only be used for specific purposes. In the UK, the Land Registry has data for public use in UK, but this does not provide the detail of the datasets held by some of the high street banks and property market apps relating to house characteristics. As such, researchers tend to use repeat sales approaches to control for this in recent applications such as flooding in England).

In the academic literature, most values (3 of 4) were obtained through administrative data (Appendix Table 6-21). One academic value and one grey literature value were obtained using enriched real estate data, which was rated as high-quality. This suggests that there is an opportunity gap in valuation research to exploit enriched real estate data to value those types of culture and heritage assets which are amenable to Hedonic Price Methods, notably heritage buildings and protected areas within urban zones.

5.5.3 RP: HP: Quality Criteria Rating

All culture and heritage values using the HP method were rated medium to high quality in the Quality Criteria Rating. Half of the values in the academic literature were rated high quality, and the others medium quality. Whilst all but one value used administrative data, their datasets were large and comprehensive. Those values which received a medium-quality rating only used standard regression models, with one value failing to account for spatial covariates.

Figure 4-8 Quality Criteria Rating, Revealed Preference: Hedonic Pricing



5.5.4 RP: HP: Conclusion

The REA shows that Hedonic methods are not used widely in the cultural heritage sector. This may be because the method is not well-suited to eliciting use values from visitors to tangible cultural institutions. For instance, the Hedonic Method does not appear to be easily applicable to cultural institutions, which provide benefits to visitors, but may not affect house prices in the surrounding area.

The small number of hedonic values in the culture and heritage literature indicates that there is an opportunity gap in valuation research to exploit enriched real estate data to value those types of culture and heritage assets which are amenable to Hedonic Price Methods, notably heritage buildings and protected areas within urban zones. The small number of hedonic studies from the grey literature may be related to the financial cost of using proprietary real estate data for non-academic research. Increased policy efforts to make these datasets open-source for culture and heritage research could help to overcome this gap in the future.

This method presents a unique data solution to the valuation of some types of historical assets. Particularly when valuing a heritage area, such as a town centre, where numerous historical assets of a certain area impact the surrounding house prices, which can be obtained by large administrative and real estate datasets that are publicly available online. This method can achieve robust valuations of heritage sites, provided that the design and datasets available are of high-quality.

In terms of methodological considerations, the majority of hedonic values were obtained through primary (i.e., administrative) data, however all these

values were marked as high-quality because these datasets were large and comprehensive. For example, one dataset included all historic properties in the neighbourhood under valuation and all properties in comparable neighbourhoods. This gives greater confidence in the robustness of this method as applied to the cultural heritage sector.

5.6 Benefit Transfer

Benefit, or Value, Transfer, is the method of transferring values from one site to another. Values can be obtained from the literature using source studies (or source sites) rather than costly primary data collection. Source site values can include use and non-use values obtained from SP and RP studies.

BT: Conclusion Preview:

Produced for a range of cultural and heritage assets, BT is a versatile method that can be designed to match the context of any culture or heritage asset type by transposing 'primary' research valuation estimates from one site to another. BT studies were more often found in the grey literature. This is likely due to the method being fast and cost-effective, and often commissioned to meet policy demands without needing new primary research. All BT studies were of a medium-high quality rating with estimates from SP survey design. Seven BT values were based on fewer than 4 study sites, which presents a concern as robust transfer testing can only be performed on a minimum of 4 study sites. Overall, we recommend the BT method in the cultural heritage sector, as it presents a fast and cost-effective way to estimate non-market values for culture and heritage assets, provided that best practice rules of minimum study numbers and transfer testing are applied

5.6.1 BT: Thematic split

The REA identified 16 values for culture and heritage using BT. Of those, around half (8) were from the grey literature. The most common reason for low quality rating was having a low number of study sites (under 3) which draws into question the robustness of the transfer error testing (since as a minimum it is necessary to survey 3 study sites to perform transfer tests, see Lawton et al. for more detail.[75]

[75] Lawton et al. 2018

The vast majority (15 of 16) of BT values were based on an underlying CV methodology, with the exception being the academic study which was based on a mix of SP and RP methods (full results in Appendix Table 6-24).

The majority of papers using BT methods valued built heritage (9, of which 2 rated medium-high quality), specifically castles (n=6) and other types of built heritage (n=3).

Of the remaining BT values, 2 were for museums, 2 for protected city areas and 2 for cathedrals, all of medium to high quality.

Built heritage: The REA identified 2 BT values as an annual tax to preserve the built heritage interiors of stately homes in Europe from climate change damages in the UK (£36.27, based on 3 study sites), and Germany (£28.05, based on 4 study sites). Both combined use and non-use values from the general public.[76]

Protected areas: 2 BT values for historic city cores, eliciting a one-off donation to reduce damage, improve maintenance and conservation in the face of the increased risks of climate change, with one value for those who had visited the city in the past 3 years (defined as a use value, of £4.27 based on 4 study sites) and one value for those who had not visited in the wider national population (a non-use value of £2.72, based on 4 study sites).[77]

Museums: 2 BT values for regional museums, with visitor WTP an entry fee (defined as a use value, of £7.29 based on 4 study sites) and WTP to keep the museum open to the general public elicited from those who had not visited in the wider national population (a non-use value of £3.63, based on 4 study sites).[78]

Cathedrals: 2 BT values for cathedrals, eliciting a one-off donation to reduce damage, improve maintenance and conservation in the face of the increased risks of climate change, with one value for those who had visited the cathedral in the past 3 years (defined as a use value, of £3.29 based on 4 study sites) and one value for those who had not visited in the wider national population (a non-use value of £1.66, based on 4 study sites).[79]

5.6.2 BT: Methodology

The credibility of BT values largely rely on the original source studies the values were obtained from (whether obtained from the literature or by primary research). As such, the majority of methodological considerations to be

[76] Mourato et al. 2014
[77] Lawton et al. 2018
[78] D. Fujiwara et al. 2018
[79] Lawton et al. 2018

accounted for pertain to SP or RP methods. Once these issues are accounted for, the BT methods (such as the number and suitability of source sites) and transfer error testing needs to be reviewed.

Errors can occur in BT due to improper inferences made regarding welfare effects and misguided policy decisions. Errors may include miscalculations transferred from the original studies (i.e., measurement errors), errors from the inaccurate or inappropriate selection of source sites (i.e., generalisation errors), and errors generated by transferring primary values to the study site in question (i.e., generalisation and transfer errors). To reduce the likelihood of such errors, a larger group of source sites is recommended as the larger the set of study sites, the lower the risk of errors.

5.6.2.1 BT: Data collection

In all but one case, BT values for culture and heritage assets were based on primary data collection (Appendix Table 6-25). The benefit of this approach is that the data collection can be designed in a standardised way that enables full testing of transfer errors (see next section). The exception was a single study in the academic literature, which was obtained for an archaeological asset (Petroglyph National Monument, USA) using meta-review of data obtained from the literature.

5.6.2.2 BT: Number of study sites

Study sites should be selected based on similar characteristics to the valuation site. For example, a source study valuing a large national theatre should not be transferred to a smaller regional theatre for risk of overestimation of value. It is generally recommended that at least 4 source sites should be used to transfer the value onto the valuation site. Anything less introduces the likelihood that the value cannot reliably be transferred. To test whether the predicted value (sourced from the study sites) applied to the valuation site is accurate, transfer testing can be conducted. The literature recommends transfer errors below 40%.[80] Anything above this cut-off cannot be assumed to be an accurate value of the site.

The Quality Criteria Rating set the cut off between low and medium quality at 3 sites because as a general rule, the larger the set of study sites, the lower the risk of measurement error related to the possible selection of a single inaccurate or inappropriate source study, with a general 'rule of three' minimum, or 4 for the purposes of running full transfer error tests.

Just over half (n=9) of the BT values were based on 3 or more study sites, of which 8 were rated medium-high quality (Appendix Table 6-26). Of those, 7 BT

[80] Brouwer and Navrud 2015

values were based on 4 study sites or more, enabling them to perform full transfer testing and giving greater confidence in the robustness of the results.

In the academic study of Petroglyph National Monument, only 3 source studies were used, of which the authors did not comment on the quality of these studies, leading to our concern that the source sites may have been unsuitable for the source site. Something of a concern is the differences between the source studies, discussed further below. For one, the academic value assumed petroglyph (i.e., rock art) sites are homogenous. While all the source studies valued preservation of petroglyphs (i.e., rock art), these sites were on a mixture of private and public land. The methods to obtain use values, and only in some studies non-use values, were a mixture of SP and RP methods.

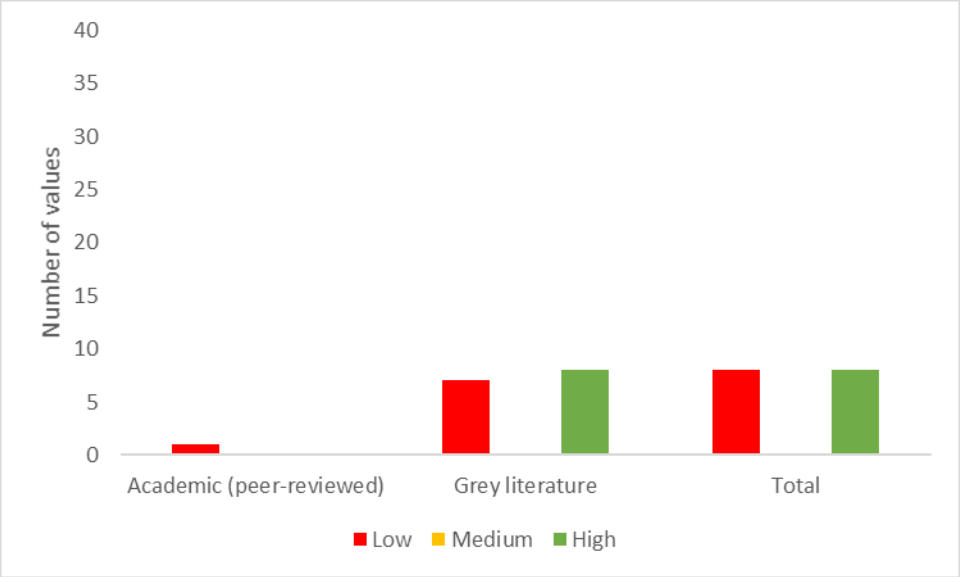
5.6.2.3 BT: Transfer testing

In the grey literature, where the vast majority of BT values for culture and heritage assets exist, all but one of the values applied the full suite of Unit, Adjusted, and Function testing (Appendix Table 6-27). However, this statistic should be interpreted alongside the other methodological considerations outlined above. In particular, the fact that 7 of these values were based on fewer than 4 study sites.

5.6.3 BT: Quality Criteria Rating

Overall, there is evidence that BT is being applied to culture and heritage assets using good methodological practice in at least half of cases. Of those rated low quality by the Quality Criteria Rating, this was most commonly due to too small a number of study sites. We recommend that further guidance is issued to the sector that as a minimum BT requires three study sites to overcome measurement error and outlier bias, and a minimum of four sites to perform the full set of transfer error tests in a robust way.

Figure 4-9 Quality Criteria Rating, Benefit Transfer



5.6.4 BT: Conclusion

BT values have been produced for a range of cultural institutions and heritage assets. This demonstrates the versatility of the method, which can be designed to match the context of any culture or heritage asset type.

The vast majority of BT studies were provided in the grey literature. This may reflect the fact that BT is commonly commissioned to respond to policy demand, rather than as research for its own sake. BT involves transposing 'primary' research valuation estimates from one site to another in a fast and cost-effective way as it means the valuation estimates can be used in other contexts. As a consequence, it may be that BT is more likely to be commissioned by government agencies than academic research funders, with the purpose of providing values which can be transferred without the need for new primary research. Academic researchers may be more inclined to research primary studies rather than needing to apply existing values to a policy site. Indeed, the language of BT (policy and study site) is indicative of the policy-focused nature of the technique, which we suspect accounts for the higher number of grey literature BT studies in culture and heritage valuation.

In terms of methodological considerations, all of the BT values with a medium-high quality rating were based on bespoke SP survey design, which enables values to be constructed which meet the needs of policy analysts, while providing a sufficient number of study sites to both ensure that the values represent a robust average of multiple sites (overcoming outlier bias) and enabling transfer tests to prevent transfer of values that would introduce high transfer error.

Of more concern, 7 BT values were based on fewer than 4 study sites. Robust transfer testing can only be performed on at least 4 study sites. We therefore urge caution in applying BT values based on only 3 study sites to business case analysis, as they will not have been subject to the full suite of transfer tests.

We recommend that BT continue to be pursued in the cultural heritage sector as a fast and cost-effective way to estimate non-market values for culture and heritage assets, provided that best practice rules of minimum study number and transfer testing are applied, in line with Johnston et al. in their Guide to Researchers on Benefit Transfer [81].

5.7 Wellbeing Valuation

WV estimates compensating and equivalent measures of welfare change from data on people's SWB. Essentially, the impact of an outcome (engagement with culture and heritage) on a person's subjective wellbeing (SWB) is estimated, for people who have actually experienced this outcome in their life. This impact is then converted into a monetary amount, by estimating the amount of income that would result in an equivalent change in SWB. It offers a robust method to estimate the value that individuals gain from engagement or participation with culture and heritage without needing to ask people directly for their willingness to pay (WTP) which, as discussed in previous section, can lead to biased responses.

In theory, any measure of subjective wellbeing can be used. SWB has three broad categories; evaluative, affective, and eudemonic.[82] Evaluative SWB measures refer to people's global assessments of their life or domains of their life. The most prominent measure is satisfaction with life which refers to people's global assessments of their life or domains of their life.[83]

WV is a relatively new method of estimating direct and indirect use values, including option values, dependent on the model that is employed. While WV does not commonly identify non-use values, appropriate survey and study design can be employed to determine non-use values of a site. WV is not recommended for sites which are not frequented often (e.g., one-off or infrequent events) and therefore are unlikely to have a measurable impact on commonly-used measures of SWB. As such, WV is suited more to sites that are regularly visited, where site visitation is more likely to have a measurable effect.

[81] R. Johnston et al. 2015
[82] Kahneman et al. 2003
[83] Daniel Fujiwara and Dolan 2015; Clark et al. 2018

Whilst wellbeing questions do not suffer from some of the 'focusing' biases SP questions do, they are susceptible to some biases of their own, such as framing effects (e.g., leading questions), order effects, time of interview (e.g., time of year), and context bias (e.g., weather, interview environment, pleasant experiences beforehand). Recollections can further be biased from intense emotional peaks and the end of the experience (i.e., peak-end rule). Research is ongoing to investigate and minimise the effect of such biases.

WV: Conclusion Preview:

More common in the grey than the academic literature, WV lacked high-quality valuation studies of culture and heritage assets. Due to the regularity of the WV method, where frequent experiences are valued to determine the impact on SWB, this method relies on assets that are visited or experienced on a regular basis. The most common asset valued in this REA was engagement with the arts, followed by visits to cultural institutions and historical amenities. This WV approach is promising, provided that the sites are frequented regularly and have an impact on life satisfaction, with the potential for high-quality non-market valuations if studies make use of the recent methodological improvements in this method (e.g., appropriately addressing endogeneity issues in modelling when estimating impact of income on wellbeing). We expect to see WV become a more popular method, particularly in the academic literature, in the near future.

5.7.1 WV: Thematic split

While less frequent in the literature than the other valuation methodologies discussed, WV has been used to value certain cultural and heritage goods resulting in a small number of values (n=13) graded in this study, coming from only 6 distinct studies (full results Appendix Table 6-28). None of the values were graded as high-quality which reflects both the fact that WV is a relatively new approach that is still being developed, and that most of the relevant studies graded were carried out in 2015 or earlier. There are more recent WV studies, outside of the realm of culture and heritage, that would be graded as high-quality on account of recent methodological improvements in the approach. This indicates potential for WV studies to value cultural and heritage goods going forward. Only a single relevant academic value was found (rated as low quality), with the rest in the grey literature (all rated as medium quality).

The asset valued most frequently in different studies using the WV approach is engagement with the arts. As the arts is often a non-market good that is free at the point of use, other valuation methods that rely on market prices or stated willingness-to-pay may not be appropriate. Closely related are values associated with visiting cultural institutions such as museums and libraries. Finally, one study contains five values of various historical amenities.

Engagement with the arts: There was one value for engagement with the arts in general in Australia (rated as medium quality) with a monetary wellbeing impact per person per annum of £2,304.[84] Another value (rated medium quality) valued participation in cultural activities in Canada implying a monetary wellbeing impact per person per annum of £5,334.[85] Two values from the same study (both rated medium) valued being an audience to the arts at £2,380 per person per year and participating in the arts at £1,744 per person per year. Finally, the academic study, rated as low quality, valued the belief of being part of a community that supports arts and culture at £21 per person equivalent to a one-off WTP.[86]

Cultural institutions: There was a value (rated as medium) that valued the monetary wellbeing impacts of visiting museums regularly at £3,753 per person per year.[87] Another value from a different study (rated as medium quality), valued the wellbeing impact of using libraries frequently at £1,498 per person per year.[88]

Historical amenities: One study produced five valuations of various historic amenities all rated as medium quality. These include the monetary wellbeing value of regularly visiting a town/city with historic character (£1,702 per person per annum), the value of visiting a non-religious historic building that's open to the public (£1,560 per person per annum), the value of visiting a place connected with industrial heritage, e.g., a dockyard, or a historic transport system (£1,274 per person per year), the value of visiting a historic place of worship for non-religious reasons (£1,130 per person per year), and the value of visiting an archaeological site (£985 per person per year). The author states that all of these values should be seen as upper-bound estimates.[89]

5.7.2 WV: Methodology

The primary methodological considerations when conducting WV include empirical design, sample sizes, and data sources. A high-quality study would take care in each of these areas.

5.7.2.1 WV: Empirical design

Various endogeneity issues can arise in WV when estimating the impact of income of wellbeing (e.g., life satisfaction). For example, if there are costs associated with earning more money such as more stress and reduced time

[84] Australia Council for the Arts 2015
[85] Lemyre et al. 2018
[86] Saz-Salazar et al. 2017
[87] Fujiwara 2013
[88] Fujiwara et al. 2014
[89] D. Fujiwara, Cornwall, et al. 2014

for leisure pursuits, then a failure to control for these factors will lead to a downward biased estimate of the impact of income on wellbeing and overly large valuations of non-market goods. Indeed, initial studies derived large, implausible values for non-market goods on account of endogeneity issues.[90]

A number of solutions have been offered in the literature. Firstly, one can make use of panel data (following specific individuals over time) to utilise a fixed effects model that controls for any factor that varies between individuals but is fixed over time (e.g., genetics). It is worth noting however that such an approach cannot control for factors that vary over time. Another approach is to make use of an instrumental variable that is correlated with income but doesn't suffer from its endogeneity problems. Use of such an approach has been shown to increase the estimated impact of income on wellbeing, resulting in smaller and more realistic monetary valuations. Such techniques can be used not only to robustly estimate the impact of income on SWB, but also the non-market good being valued on SWB.

Other advanced empirical designs may involve integration with SP or Geographic Information System (GIS) data. In the UK, for instance, it may be possible to combine data on individual wellbeing from national household surveys like the Annual Population Survey and Understanding Society with GIS data on the presence of built and industrial heritage collected by Historic England, to analyse the wellbeing impacts of heritage.

As shown in Appendix Table 6-29, most of the studies graded do make an attempt to deal with the endogeneity issue when estimating the impact of income on SWB, although solely through the use of an instrumental variable for income. There is a distinct lack of studies that have utilised a fixed effects model, either for the estimation of the effect of income on SWB or the effect of the good being valued on SWB. As such there appears to be potential for studies to exploit the availability of panel data to a greater extent to provide more robust valuations of cultural and heritage assets.

5.7.2.2 WV: Sample group

WV should not be used to value one-off or infrequent events. This is because these events are less likely to hold a large effect on an individual's life satisfaction compared to regular engagements, such as regular visits to a museum.

For the majority of WV values graded in this study, we cannot be certain that all respondents enjoy regular engagement with the good being valued. For example, some values are based on those who have one or more visit to the site in question in a given period (Appendix Table 6-30). Whilst some of these people may visit the site regularly, others may have only visited once

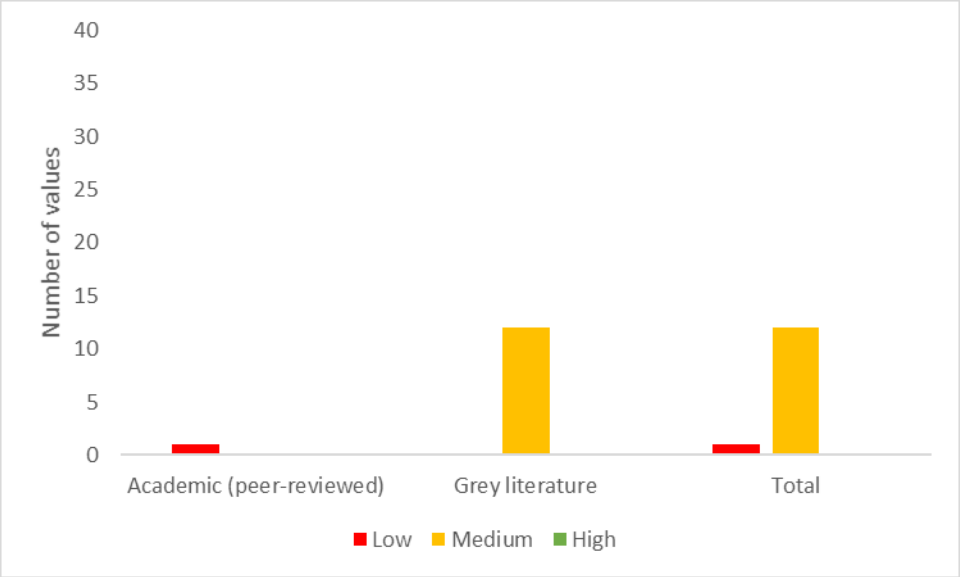
[90] D. Fujiwara et al. 2012

and so may not reflect an increase in measured SWB. In fact, only one study (accounting for three of the values) explicitly takes into account the time spent visiting the site in question (museums). Once again it appears that there is potential for higher quality WV studies to be carried out in the cultural/heritage space that ensure the sample has regular engagement with good being valued.

5.7.3 WV: Quality Criteria Rating

Overall, one value was graded as low quality and the rest of medium quality. Whilst studies generally made use of suitable datasets with large sample sizes, there are improvements that can be made in terms of empirical design and the choice of respondents to include in the analysis. Specifically, WV studies could exploit panel data further by making use of fixed effects models to more effectively deal with endogeneity problems, and could even look to integrate with GIS data. Studies could also take more care to ensure that respondents regularly engage with the good being valued so that any wellbeing impacts can be captured in SWB measures such as life satisfaction.

Figure 4-10 Quality Criteria Rating, Wellbeing Valuation



5.7.4 WV: Conclusion

Similar to the BT method, WV is not a commonly used method in the academic literature, although the REA shows that it is more common in the grey literature.

However, the REA finds a lack of high-quality WV studies in the area of culture and heritage which indicates potential for further WV studies in this area that make use of recent methodological improvements in the method.

The nature of the WV method leads it to valuing regular experiences that may have a measurable impact on SWB. As such this REA found the most common cultural and heritage asset to be valued by WV to be engagement with the arts. This was followed by visiting cultural institutions and visiting certain historical amenities.

WV presents a promising approach in non-market valuation, provided the valuation sites are frequented often and so can have an impact on the life satisfaction of the surveyed population. We expect there to be a dramatic increase in the number of WVs published in the academic literature in the coming decade as this method becomes more popular.

In terms of methodological considerations, care should be taken over potential endogeneity issues when estimating the impact of income on wellbeing so as not to overestimate the values for non-market goods. Limitations in the WV studies graded in this study imply potential for improved WV studies to be carried out in the future to value cultural and heritage assets.

6 Conclusion and next steps

6.1 Overview

This REA, conducted on behalf of DCMS, identified values elicited using six best-practice non-market valuation methods for a range of arts and cultural assets. It applied novel data science techniques to collect 171 relevant, academic papers and grey literature reports (Government and third sector publications) on valuation of culture and heritage assets.[91]

The Quality Grading Criteria developed for this REA has never been attempted before in the cultural or environmental sector, and its synthesis of best practice guidance[92] and direct researcher experience provides a quality RAG rating for all of the values in the Evidence bank, representing a considerable contribution to the sector.

The REA results suggest that established non-market valuation methods, using SP and RP methods, are more prevalent in the academic literature than the more recent and novel methods like WV and BT. In the grey literature, CV

[91] We acknowledge that the final shortlist of papers were largely determined by online algorithms. Algorithms on online databases, such as the widely used Google Scholar, determine paper relevancy and rank them according to pre-determined features (e.g. number of citations). Therefore, any papers that were ranked lower than this cut-off would have been excluded based upon our cut-off points. While we attempted to expand the number of search hits we assessed (e.g. the first 1000 search hits for Google Scholar), there is no guarantee we were able to find all the papers that have been published in this field. Particularly if these papers did not meet the standards set-out by the algorithm deemed to be highly relevant and high in ranking. Any future research should keep this in mind when conducting reviews of the literature.

[92] G Atkinson et al. 2018; Bateman et al. 2002; R. J. Johnston et al. 2017

(either single studies or as part of BT) are also the most common, but there is also a rising number of WV studies. We hypothesize that this may be driven partly by policy need, since CV and WV provide valuations for the cultural heritage asset as a whole, rather than attributes of the asset (as in DCE) and this is more relevant for government business cases and funding bids.

Overall, there is evidence of an inconsistent empirical record that is well represented for certain types of tangible culture and heritage asset, such as museums/galleries/theatres and libraries and built heritage sites. Less well represented in the valuation literature are digital services, industrial heritage, and religious assets, which we outline in the Research Gaps section below. There is also a narrow range of non-market valuation methods used, in particular CV and DCEs, although there may be some evidence of evolving trends in the enhanced use of WV, which uses observational data estimate the value that individuals gain from engagement or participation with culture and heritage, and BT, which involves transposing 'primary' research valuation estimates from one site to another in a fast and cost-effective way as it means the valuation estimates can be used in other contexts.

The aim of this REA was to establish what cultural and heritage assets had been valued and the economic methods employed. To define best-practice, Quality Grading Criteria was established to grade the literature methodology, thereby creating straightforward evidence bank. This evidence bank contains the values found within the literature. It was not within our scope to conduct a statistical analysis on these values (i.e., a Meta-Analysis). The Quality Grading Criteria was used to evaluate the methodological quality of each value. Thereby, values from the same study may have differing grades. Future analysis should consider the extent to which non-market values in this REA and existing values used in other sectors can be combined in an additive way. It is outside of the scope of this study to consider such double counting issues in detail. The probability is that welfare-based valuation will capture more of the surplus values than cost replacement methods, but it is important to be aware that there may be some additional issues to be considered as further conceptual work is undertaken around culture and heritage stocks and services and how these align with existing natural capital approach to valuing natural environments.

6.2 Research gaps

Our hope is that this review will help to identify gaps, enable the prioritisation of research streams, and identify ongoing methodological challenges. The largest research gap regards the lack of Digital cultural heritage assets and services (excluding Public Service Broadcasting) valued using non-market valuation methods in the academic literature to date. At first, this may reflect the relatively recent emergence of digital assets and services in the cultural and heritage sectors. However, we note that all the values for digital assets related to public service broadcasting. This may reflect the policy need which public service broadcasters often find in

having to justify their public funding in non-market valuation terms, and that the vast majority of reports come from the grey literature reflects the interest from public service broadcasting in commissioning non-market valuation research to increase the evidence base in this area. The risk, in such cases, is that the valuations lose their independence due to reliance on non-academic valuations, rather than a healthy mix of both, and one cultural good commanding the digital domain. For this reason, it is important, as we do, to distinguish between those values that apply good practice methodologies to minimise bias.

The absence of values for other types of digital assets is surprising, given the recent emergence of online portals for digital cultural services in many countries (such as virtual tours of art galleries and the digitation of cultural archives (such as the British Film Institute Unlocking Film Heritage project[93]). This shift to digital services has not yet been reflected in the non-market valuation literature, either in the academic or the more responsive grey literature fields. We anticipate that this may experience a step change in the coming decade, especially in response to the 2020 COVID-19 pandemic which led to the closure of many cultural institutions for many months and an acceleration of their digital offer.

Already, United Kingdom cultural institutions have responded to the digital trend during the lockdown period. Not to forget the numerous operas and theatre productions which are widely available on platforms such as Youtube and BBC iPlayer. British Museum offers a digital walk-through of a hand-picked sample of their collection with curator talks. Manchester Museum grouped their digital content into one virtual exhibition with curator tours of special exhibits. London's Street Museum of Art, together with Google, are presenting a virtual street art tour around Hackney and Shoreditch. With the Natural History Museum, Birmingham Museum and Art Gallery, National Museum of Scotland, The Ashmolean, Jorvik Viking centre, National Museums Liverpool, and Stonehenge following this virtual trend too. This digital movement will present an exciting future opportunity in valuation research for the cultural sector.

Very few non-market values were identified for industrial heritage or religious assets, despite the fact that these types of physical built heritage asset are amenable to both SP and RP methods. This indicates an opportunity gap in valuation research to exploit enriched real estate data to value those types of culture and heritage assets which are amenable to Hedonic Price Methods, notably heritage buildings and protected areas within urban zones. The small number of hedonic studies from the grey literature may be related to the financial cost of using proprietary real estate data for non-academic research. Increased policy efforts to make these datasets open-

[93] https://www.bfi.org.uk/britain-on-film/unlocking-film-heritage

source for culture and heritage research could help to overcome this gap in the future. Where HP models have been applied to culture and heritage assets, the REA findings suggest that datasets available for hedonic analysis are strong across multiple jurisdictions. This has resulted in a healthy set of hedonic values for those cultural heritage asset types which are amendable to this method (i.e., those assets which are expected to affect local house prices). We would therefore expect this to continue into the future with increased access to online land registry and house market data. We would note that some industrial heritage, such as canals, may be valued already using cost replacement methods, or consideration of their contribution to landscape values or ecosystem services, which fall outside of the non-market valuation methods included in this REA.

There may be opportunities to use TCM in a more systematic way to value audience engagement with cultural and heritage sites using existing time-use surveys collected at the national level.[94] There may also be existing models from other sectors that could be followed to provide Travel Cost estimates of the value of visits to heritage or religious sites in the UK. For instance, Defra produced the Outdoor Recreation Valuation tool (ORVAL) to understand how people value the country's natural environments and greenspaces using the Monitoring of Engagement in Natural Environment Survey. It may be possible to combine this model with emerging data visits to cultural institutions, heritage sites and religious buildings. Even accepting that the TCM provides a partial value for the value of a trip (rather than the total use and non-use values of the site), it would still be useful to know the value of cultural resources even in those narrower terms. In other sectors, such as transport, there have been recent efforts to understand the value of a landscape more than sum of its parts, which incorporates amenity values in more of a place-making sense that aligns with recent valuation in the heritage sector (Historic England research, forthcoming).

An additional gap in the research is that we found no examples of valuation studies that applied welfare weighting to the values estimated for culture and heritage assets. Welfare weighting, as recommended by HM Treasury Green Book, permits using distributional weights to adjust for diminishing marginal utility of income. We found no instances in which welfare weighting was pursued in the culture and heritage REA. Consequently, non-market valuation in the culture and heritage sector does not account for how inability to pay might constrain stated willingness to pay values, possibly leading to inflated willingness to pay values for sites frequented by higher socioeconomic users compared to sites which are frequented less by such users. A distorted funding decision may arise as a potential result of this, as sites with higher socioeconomic visitors, and thereby greater aggregated value, might access more funding than sites without this audience. We

[94] https://data.gov.uk/dataset/ba226a1a-de1e-4f9e-80e6-835e3518764e/time-use-surveys-and-the-measurement-of-national-well-being

recommend that future applications of non-market valuation methods consider the application of welfare weighting in instances where there may be significant differences between the values held by gainers and losers in the population.

6.3 Comparison to the Environmental sectors

This REA makes an important contribution by establishing where the arts and cultural sector are, and can be usefully compared with those from other (policy) sectors such as environment. Environmental valuation has been a vibrant sub-field of environmental economics and environmental studies over a number of decades resulting in a deep empirical record. This record is also extremely broad given the disparate policy areas to which environmental valuation has been applied, such as waste impact on natural resources to environmental biodiversity protection.

A more coherent and focused basis for comparison is arguably the subset of environmental valuation represented by studies which value ecosystem services. These values are generated by ecosystem assets, which, analogously with (some) cultural assets, are combinatorial in the sense that they comprise of component resources that together can be conceived as an identifiable broader asset.[95] Ecosystem services themselves refer to a wide range of impacts on human wellbeing. That is, ecosystem services variously can be inputs to economic production (of businesses and so on) as well as inputs to household wellbeing.[96] Another way of looking at ecosystem services is in terms of the function they serve (e.g., providing foodstuffs and so on). Conversely, cultural assets are typically only one crucial element of those functions.

There have been numerous ongoing efforts to systematise and cohere the empirical record on ecosystem service valuation. For example, the Ecosystem Services Partnership maintains a database, building on an earlier international review by TEEB[97] and currently comprises more than 600 studies.[98] The Defra database, supporting its *Enabling a Natural Capital Approach*, collates more than 100 UK sources of valuation evidence.[99] The empirical record summarised in such databases is illustrated more generally by meta-

95 For example, a number of cultural assets such as art galleries and museums can be thought of assets while comprising a number of distinct sub-assets (artefacts, the structures within which these are housed and so on). Broad habitats such as coastal margins and moorland are made up of biotic and abiotic resources along with natural processes. This can be contrasted with natural assets which are single resources: e.g. sub-soil assets.

96 As in the case of cultural resources, this pathway to household wellbeing in the case of ecosystem services can be of two main types: indirectly where used as inputs – along with the purchase of market goods – to the production of some economic good or service such as a recreational experience; or consumed directly in generating wellbeing without any (intermediate) household consumption where these services are inputs such as 'non-use' or 'passive-use'.

97 The Economics of Ecosystems and Biodiversity.

98 Source: https://www.es-partnership.org/services/data-knowledge-sharing/ecosystem-service-valuation-database/ (accessed 21/05/20).

99 Source: https://data.gov.uk/dataset/3930b9ca-26c3-489f-900f-6b9eec2602c6/enabling-a-natural-capital-approach (accessed 21/05/20).

studies in the published literature. Two examples are discussed briefly below.

De Groot *et al.*[100] draws on 320 publications (across 300 case study locations) of ecosystem service values across the world. In total, this provides 665 specific estimates of value. Of these, more than half are direct market valuations and a further fifth are cost-based approaches (including replacement and restoration costs). This leaves relatively few (less than a third) estimates which draw on the more commonly accepted non-market valuation methods. The majority of these are SP studies, followed by Travel Cost approaches and a relatively small handful of hedonic studies.[101] Another way of categorising this empirical record is by specific ecosystem services. Notably, cultural services make-up just over a fifth of the estimates in this meta-study. Of these the vast majority (more than four fifths) were estimates reflecting recreational demand.

A more recent study is Reynaud and Lanzanova (2019) which focuses on ecosystem service values arising from lakes. This draws on 19 meta-studies published between 1999 and 2016 from across the world, with notable findings. First, in terms of cultural values, most of the underlying data comprise some form of recreational value. Although a number of these studies do contain values which reflect broader aesthetic and symbolic appreciation of lake ecosystems. Secondly, the empirical record draws on a wider set of valuation methods including TCM, HP, and DCEs. Thirdly, a larger proportion of studies are still concentrated in certain parts of the world (notably North America and Europe) and the empirical record reflects valuation traditions in those locations. Lastly, and nevertheless, the evidence base is generally expanding; of the 8 meta-studies published before 2010, 37 studies on average were reported, whereas 11 meta-studies published since 2010 reported 99 studies on average.

This brief and partial review provides some contrasts and similarities with the evidence on cultural values discussed elsewhere in this report. The empirical record on ecosystem services is larger and more diverse reflecting the large financial and human resources that have been directed to this over decades of research across related disciplines (economics and environmental science). However, there are trends which seem to align with findings for values of cultural resources. While ecosystem service valuation studies are voluminous, the evidence base for cultural services is relatively small and while a range of methods is used there is it seems a relative preponderance of SP studies (notably CV). Moreover, none of the above studies explicitly

[100] de Groot et al. 2012

101 A number of further studies are described by Groot et al. (2012) as "other".

mention WV as part of the empirical record. However, as discussed elsewhere in this report, this we expect to change in due course.

In this respect, Atkinson *et al*. (2018) provides some narratives[102] of the evolving use of non-market valuation in UK environmental policy. This is a useful complement to reflections and inferences drawing directly from the empirical record. A few findings and trends are worth noting in this respect. Initial original studies reflected a premium on gathering primary data in areas of policy concern but as the empirical record grew, along perhaps with budgetary and time constraints and a wish to boost ease of use, there has been an increased emphasis on secondary data sources (such as databases). There have been trends in relative prominence of methods in distinct areas of environmental policy. Early studies for policy use tended to utilise SP methods and this 'tradition' has continued notably in the context of the water environment (i.e., Environment Agency and water companies). However, in the context of natural capital valuation, there has been a noticeable emphasis on RP approaches and the use of ecosystem and land-use models in conjunction with environmental valuation.

102 Drawing on documentary evidence and qualitative interviews with policy officials.

7 Appendix

7.1 Literature review: Previous meta-review of culture and heritage valuation studies

In 2003, Noonan et al.'s review focused on CV studies in the cultural sector, including the arts, historical sites, theatre, museums, heritage, archaeological sites, broadcasting, libraries, and sports sites. They provide a summary of selected published CV studies reporting the survey size, cultural good in question, and the average willingness-to-pay of respondents listed chronologically, geographically, and by topic. While detailed in review, the Noonan paper makes limited attempts to assess the quality of the studies reviewed only noting that dichotomous choice formats in preference-based methods are the preferred elicitation mechanism. Beyond this, they state that "a full description of the different question formats and their effects on estimation is beyond the scope of this paper".

Similarly, Provins et al.[103] (2008) reviewed the economic values associated with historic built heritage. Their review included both SP and RP studies (but no WV) and categorised the studies by the type of good valued, grouped by geography. The review included single heritage sites (like cathedrals, castles, or towers), or groups of historic buildings (like city centres, groups of monasteries, and archaeological sites). Conclusions are made on the state of the valuation of culture and heritage at the time of writing, noting that the total number of studies reviewed (30) is decidedly small, particularly in relation to the comparable valuation of environmental goods, where the number of published studies runs into the thousands. The authors note that while use unit values are typically higher than non-use unit values, non-use value may account for a larger proportion of the whole population's value for the site. Consequently, the aggregated non-use value could be implied to account for a substantial percentage of the Total Economic Value (TEV) of the historic site, though this percentage is inconsistent across studies.

The authors note that the existing body of literature varies widely in terms of the heritage assets considered and the nature of and the way in which benefits are evaluated such that "the quality and reporting of studies and estimated economic values vary greatly". Specifically, valuation studies differ in terms of:

i. The heritage asset, that is, what it is and where it is;
ii. The historical and cultural associations of the asset (which typically make it important and often unique);

[103] Provins et al. 2008

iii. The context of the change being valued, which may be explicit in terms of incremental improvements in quality (e.g., renovation works) or more implicit (in the case of valuing access over loss of access);
iv. The characteristics of the relevant population (i.e., visitor/user vs non-user, local vs national, individual vs household payment, all of great relevance for interpreting and comparing values, and in the process of aggregation for business case analysis; and
v. The valuation methodology (for instance Travel Cost approaches will not capture non-use values).

Finally, Willis (2014) attempted more of a quality assessment of studies which applied SP Methods to value culture and heritage. However, this was not a comprehensive review and was confined to culture and heritage using SP methods only.[104]

7.2 REA Methodology

7.2.1 Stage 1: Initial Scoping

Initial Scoping determined the key search terms that were employed for the REA. Initial scoping employed Google Scholar to estimate how many hits would be returned for each search. All searches were designed to follow the formula:

(culture or heritage asset, with or without clarification terms) + (primary research valuation method) + (valuation method clarification terms)

Key terms were signified with quotation marks to return only exact terms found anywhere within each paper (i.e., within the title, abstract, main text, key words, or references). Thereby, only papers that included all of the key terms in the paper would be returned by the online databases. Boolean operators (e.g., AND/+, OR, -/WITHOUT) were used to expand and reduce the search scope where necessary. For example, "castle" ("heritage" OR "historic" OR "protected") + "Contingent Valuation" + "stated preference" + "accept". Some online databases, such as Google Scholar, have a search algorithm that favours highly cited research[105]. It was therefore decided to include the first 1,000 returns of each search to avoid ranking bias.

Umbrella terms were excluded as these terms returned too many results that could be reviewed within the short timespan, of which a large portion were

[104] K.G. Willis 2014
[105] Beel and Gipp 2009

irrelevant to the study's aim and criteria. These terms included: art or arts, heritage, historic, culture, value, benefit, cultur*, cultur* facilit*, historic* interest, public, environment, literature, and willingness. For example, the inclusion of the key term "*literature*" returns all papers that include the word anywhere within the paper, even if only commenting on the respective field's literature.

In the nature of REA, some papers may have slipped through the search net during our rapid assessment. As such, key studies to be included in the shortlist were identified *a priori* by searching reference lists of previous literature reviews in the economic valuation of the culture and heritage sectors and paper recommendations by stakeholders. We believe that our assets-based terms approach, with primary valuation method and clarification terms, returned a comprehensive list of research papers. Initial scoping searches revealed that the inclusion of umbrella terms included more irrelevant material, which would have resulted in more time dedicated to screening out these irrelevant papers. Furthermore, articles that do not list the physical asset being valued (e.g., theatre) but include a more generic term (e.g., culture), arguably would not meet the grading criteria to be included, as our grading criteria requires the physical asset to be described and named.

7.2.1.1 Cultural and Heritage Definitions

Definitions were developed through detailed review of previous sector research by Historic England (HE) Listings[106], the Department for Digital, Culture, Media and Sport (DCMS)[107&108], and the Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA)[109].

Culture and Heritage physical assets were defined as:

- **Arts**: art galleries, public arts, street art, museums, concert or town halls, festivals, music venue or bandstands or amphitheatres, theatres or playhouses or opera houses, libraries, digital and heritage archives,
- **Historic protected culture**: areas, sites, places, spaces
- **Heritage, historic, or protected**: gardens, castles, ruins, cinemas, plaques, rail, bridges, canals, mines, quarries, warehouses, mills, factories, waterways, ports, docks, harbours, aqueducts, and more generally as built heritage

106 https://historicengland.org.uk/listing/

107 DCMS (2016). The role of culture, sport and heritage in place shaping report; https://www.gov.uk/government/publications/the-role-of-culture-sport-and-heritage-in-place-shaping

108 DCMS (2010). Measuring the value of culture: a report to the Department for Culture Media and Sport; https://www.gov.uk/government/publications/case-programme-understanding-the-drivers-impacts-and-value-of-engagement-in-culture-and-sport

109 RSA (2015) Heritage, Identity and Place Report; https://www.thersa.org/globalassets/pdfs/reports/seven-themes-from-the-heritage-index.pdf

- **Historic protected:** towns, cities, businesses, transport, roads, built, buildings, structures, parks, and landscapes
- **Historic or heritage:** cathedrals, churches, chapels, mosques, temples, synagogues, monasteries, shrines
- **Public:** monuments, sculptures, statues
- **Archaeological sites**

This initial scoping produced approximately 73,430 returned search hits for initial review. A large portion of the items were expected to be irrelevant (e.g., valuing an asset not to be included, such as cultural customs), duplications (i.e., the same paper more than once), or excluded for other reasons (e.g., not published in English). This is to be expected given the broad scope of the REA (i.e., largely worldwide with no specific population and valuing all cultural and heritage assets).

7.2.1.2 Reduction in Scope

Even with automated screening methods processing the majority of citations to be screened out, this type of literature review requires quick processing (usually within 2-3 months) of which the number of citations to be processed was not achievable. This number of citations is too large for a REA to feasibly review and synthesise within the project timeline. Therefore, the scope was reduced by:

- Restricting the assets to be valued to **physical assets only** – thereby excluding concerts and performances
- Removing **null hits** from assets (e.g., "ritual art" from assets as it returned 0 results related to economic valuation)
- Reducing the **geographical scope** (e.g., predominantly OECD countries, namely North America, Europe, Australasia and South Africa, as these countries are hypothesised to have a more common set of cultural values, and have a similar cost of living range, which reduces error when transferring values into GBP£). Whereby, countries with cultural links, shared language and religion are more likely to share similar cultural beliefs and values regardless of their proximity to one another.[110] Even so, we expect differences in the definitions of art, differences in funding for public arts, and differences in cultural taste for the arts across Western OECD countries.[111]
- For grey literature, given the large amounts of potential search terms obtained through Google, the reviewed literature was restricted to the first 20 hits per search, as these results

[110] Belot and Ederveen 2012
[111] Throsby 1994

are ranked by the algorithm with greater relevancy to the search terms.

7.2.2 Stage 2: Forming the Assessment Protocol

7.2.2.1 Determine final search criteria

After initial scoping, the final inclusion criteria were developed:

- Physical cultural and heritage assets (excluded sports and environment due to the large scope)
- Research articles published within the past 20 years (i.e., 2000-2019); key articles outside of this time scope (highly cited) will be considered for inclusion on a merit basis
- Methods included were SP CV (willingness to pay and accept), SP DCE, RP HP, RP Travel Cost, WV, and BT. These methods are deemed by the HM Treasury Green Book (2018)[112] as best practice methods for this type of valuation and contained within Atkinson et al. (2018) [113] , which represents the best welfare-consistent methods[114].

As such, the following were excluded:

- Books or book chapters as published books are typically not accessible online, which makes it difficult to include them in a short time restriction (i.e., REA). In most cases, books do not present novel primary empirical research, and commonly synthesise evidence from published papers[115].
- Patents, Indexes, Citations (i.e., not primary empirical research)
- Papers not written in English as this would require translational resources and we feel the scope will be comprehensive without including non-English papers
- Natural capital (e.g., natural landscapes) were excluded unless there is a cultural or heritage element involved (e.g., Royal parks, heritage gardens, World Heritage Sites)[116] as these would already be accounted for according to academic and DEFRA (this was determined from stakeholder interviews).
- Non-OECD countries were to be excluded with the justification that reporting SP values (i.e., willingness-to-pay) is

[112] HM Treasury 2011
[113] Giles Atkinson et al. 2018
114 The Delphi method was also included in the initial literature searches for comprehensiveness, but was not to be included in the research synthesis.
115 We note that the bank of studies developed by the REA should not be seen as comprehensive. We encourage the cultural and heritage sector to continue to build on the list with any books or reports which have been omitted but could provide useful empirical evidence.
116 This significantly increased the citations, however, the consensus was to exclude it in the final papers to be reviewed for research synthesis provided it was included in the report appendix for future research in the sector.

comparable across countries if these values are adjusted to the income differential of the country in question. Thereby, this requires a differential analysis for each country involved. Additionally, there might be cultural differences to account for, such as the inclusion of foreign values (e.g., tourist values). For simplicity, European, North America and Australasian OECD countries provide better comparability when adjusted, with minor concerns of cultural differences.

7.2.3 Stage 3: Effect Extraction and Analysis

7.2.3.1 Search methods

Multiple databases were searched to ensure sufficient coverage; Google Scholar, Science Direct, and relevant websites (e.g., Government department website, found in searches by using the Google search engine). Google Scholar provides an open-access broad database that includes a large portion of academic and grey literature. It is recommended that only the first 200-300 results returned per search in Google Scholar should be included in literature reviews due to their greater relevancy.[117] However, as our key search terms did not include umbrella terms (e.g., "cultur*") and the Google algorithm favours highly cited research, it was decided to review the first 1,000 search hits from each search to ensure papers that may have not been ranked as high due to using broader umbrella terms and which have may not have been highly cited were identified. The alternative academic online database, Science Direct, is a reputable open-access database that is considered to have greater reliability than Google Scholar. However, Science Direct contains fewer papers in total; meaning a narrower scope of literature within the database. Google searches were additionally used to obtain grey literature from government departments and industry. Search hits were restricted to the first 20 search hits, as Google searches retrieve a large amount of irrelevant material.

Searches were conducted by a researcher through Science Direct and Google Scholar between 23rd January and 9th March and returned over 85,000 combined search hits. Grey literature was scoped using Google and returned over 6,900 search hits between 20th and 21st February[118]. This brought the total number of search hits to over 91,900. Every search hit was recorded into an excel file and saved with the search terms used. Due to the number of duplications identified across our search hits, we were confident that we had reached saturation. This high number of duplicates was thought to be due to multiple databases being searched and searches broken into multiple searches. For example, the same paper could have been identified using different valuation method search terms if the terms were included anywhere in that paper. That

[117] Haddaway et al. 2015

118 When the first 100 search hits were included, Google returned over 31,000 search hits. The scope was reduced to the first 20 search hits as the large portion of the lower ranked results were deemed irrelevant.

is, we were confident we had found all the relevant papers using this method. From these search hits, automated and manual screening methods were used to exclude papers that did not meet the inclusion criteria.

7.2.3.2 Text-mining methods

While text-mining is an innovative strategy for literature reviews, it has been previously used in the literature (e.g., refer to the DCMS CASE 2010 Report[119]). To ensure we followed ethical data science practice, we consulted Office National Statistics's (ONS) best practices.[120]

The automated code is run by the software program Python (and Pycharm). This method screens article titles and abstracts on basic characteristics (language, article accessibility, irrelevant material, redundancy – patents/books, and marks duplicates) reducing the workload and time restrictions on researchers. For example, this method had approximately a 76.4% accuracy rate in accessing and saving abstracts for screening. Of the 468 articles returned, 305 papers were deemed successful, 69 were excluded, and the code failed (i.e., Failures) to obtain and code 94 papers based on a partial search (see Figure 6-1).[121]

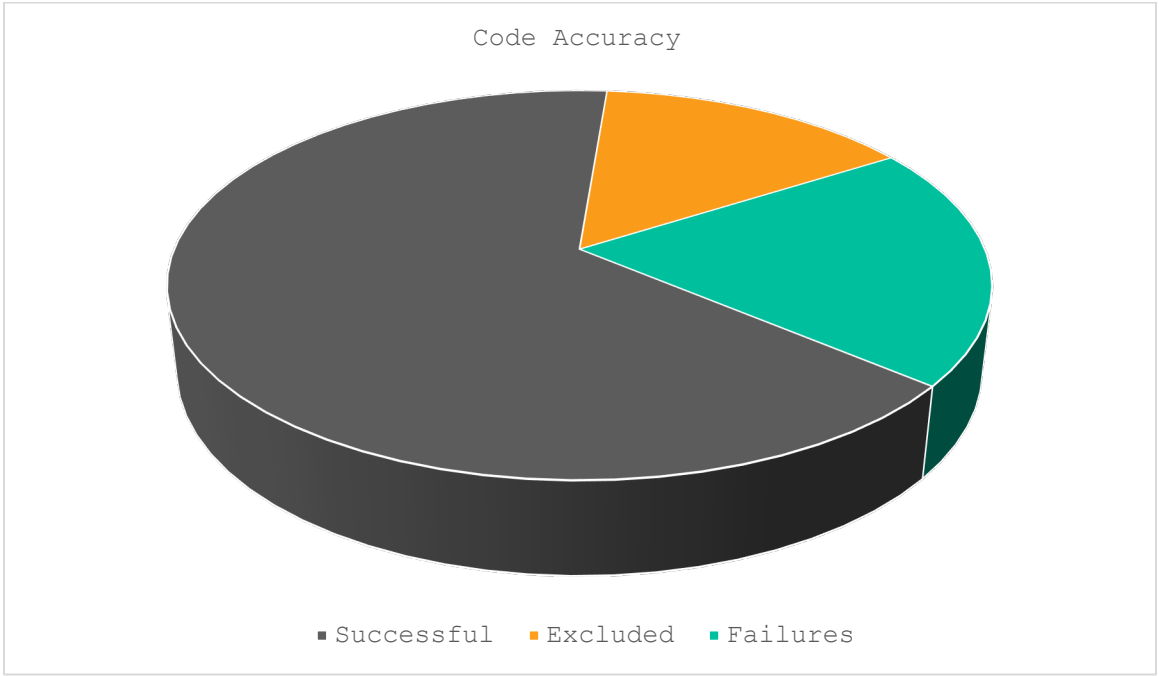


Figure 6-1 Accuracy of the Code in accessing and saving abstracts for screening

[119] Marsh et al. 2010

[120] https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/webscrapingpolicy

[121] As this is still a new method, there is no established guidance on what is an acceptable level of error when employing text-mining methods. Our research team was happy with this promising level of accuracy within our code as researchers were verifying any papers flagged by the code for exclusion.

It is important to note that in all stages of the automated methods, researchers checked what the code flagged for exclusion. Thereby, no papers were excluded without the approval of a researcher. As outlined below in Figure 6-2, text-mining methods were crucial in screening out irrelevant papers. The automated process is outlined in more detail in the following three sections.

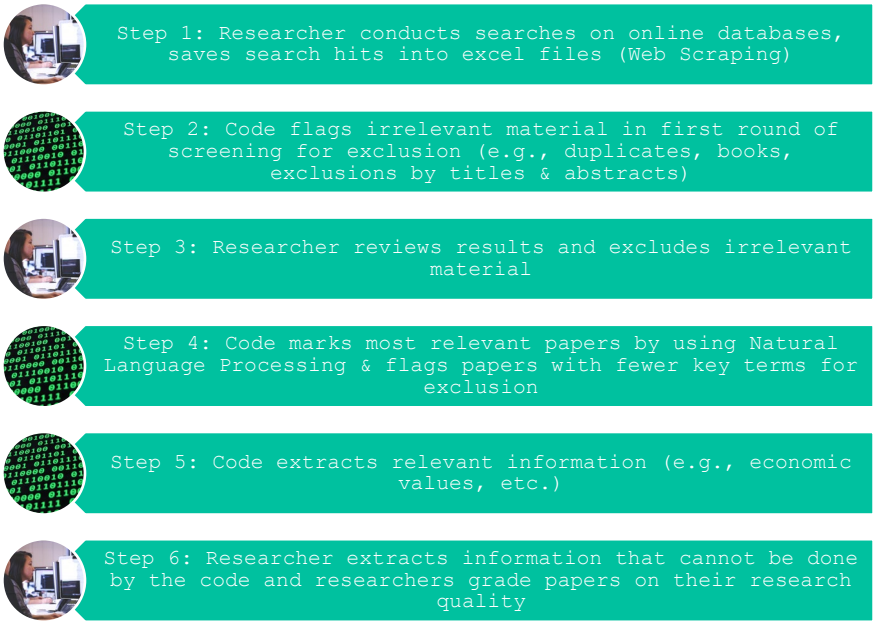


Figure 6-2 The Stage 3 process broken down into steps

Exclusions and abstract collection:

1. The code accesses and retrieves information in the .csv file which contains all the search hits of a given search (e.g., *"hall" ("public" OR "concert") "value transfer"*). Any duplicates within this file (or any files previously processed) are marked as a duplicate and discounted (i.e., screened out).
2. The code marks any search hit which falls outside the timeframe (e.g., older than 20 years) and discounts these search hits.
3. The code then searches through the title and abstract for irrelevant key terms. Irrelevant key terms were identified by a researcher based on irrelevant papers found within the search hits and programmed into the code. For example, a lot of financial papers were being returned (key terms: bitcoin, banking), which fell outside the scope of the REA. Any irrelevant key terms can only contain a maximum of two words within the key terms for the code to identify and screen these irrelevant papers out.
4. The code then identifies the language of the title and provides a confidence level on the certainty of that language. Those search hits with a log probability greater than 50 that title

of the paper is not in English are marked as Non-English and discounted.

5. The code searches for irrelevant material (patents, websites, or books including sections, chapters or indexes).

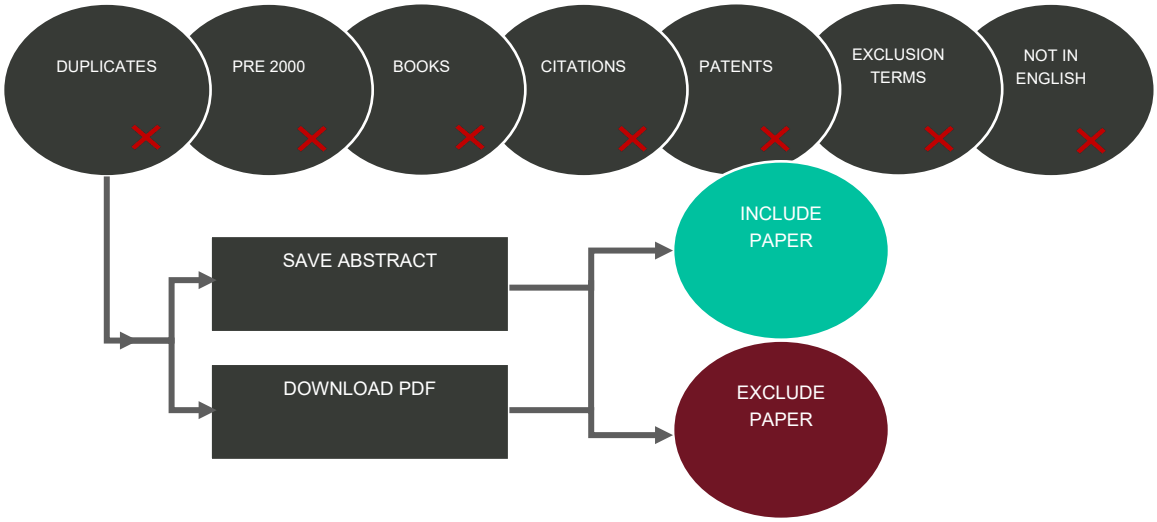


Figure 6-3 Code flags irrelevant material in first round of screening for exclusion (e.g., duplicates, books, exclusions by titles & abstracts)

6. The code then downloads the PDF or extracts the abstract from the download page. Where neither of these are possible the paper is flagged for further investigation by a researcher.

Papers marked for exclusion by the code (and by researchers) were discounted from the process, as they were non-peer reviewed or irrelevant material. Non-peer reviewed papers refers to conference or discussion papers (e.g., Working Papers), theses, or books (including book sections or chapters) that do not take part in the standard academic journal article peer-review process and were therefore excluded due to the absence of this structured vetting process. Book sections or chapters, while arguably are peer-reviewed, were excluded due to the lack of research typically contained in these papers. Irrelevant material refers to anything papers or items that slipped through the previous exclusion process. This includes papers that were not valuing arts and heritage assets that were not found by our exclusion terms in the first automated exclusion stage.

Cleaning by Researchers:

1. Researchers reviewed the papers the code was not able to screen and mark for inclusion or exclusion. Papers marked for exclusion were determined by researchers screening the relevancy of their

titles (*Inter-rater reliability*).[122] If there was any uncertainty regarding the paper for exclusion, the paper abstract was then read for screening.

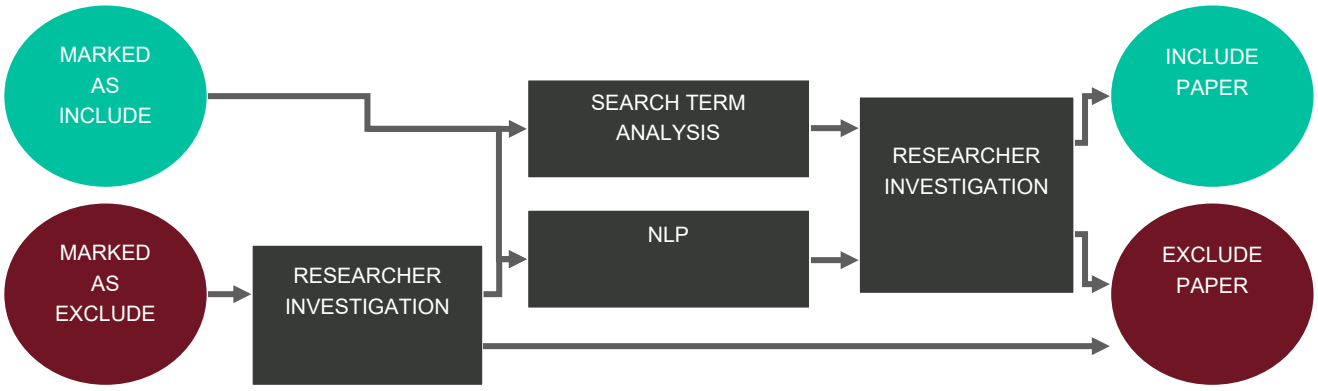
2. Any papers that were marked for inclusion but failed to save are reviewed by a researcher.

Natural Language Processing (NLP) on full text (via pdf file):

1. The code reads the pdf files by searching for our key terms (economic methods, asset type, heritage words), nouns (place or asset names), and exclusion terms[123]. The key asset terms are weighted (e.g., art gallery, theatre), as we consider them to have greater importance than the other information the code is searching for. These terms were extracted into an excel file to easily show which search terms are being found in each paper.
2. The code then marks each paper based on its relevancy. If the paper has referenced our key terms more than once, it returns a higher relevancy result (the most relevant score being 1, the least relevant score being 0). Least relevant papers can be excluded based on an arbitrary cut-off (a relevancy score of 0.5)[124].

3. Researchers then review the papers marked as highly relevant by the code and exclude irrelevant papers based on their titles and abstracts. This will ensure only highly relevant articles (regarded as highly relevant by the code and researcher) to be scanned and graded on research quality.

Figure 6-4 Researcher reviews results and excludes irrelevant material. Code marks most relevant papers by using Natural Language Processing & flags papers with fewer key terms for exclusion.



122 Inter-rater reliability was tested on a sub-set of papers (approx. 300 papers, with 97% accuracy), ensuring that moderation was sufficiently established.

123 Exclusion terms are identified by researchers in all stages from observing patterns of words and repeatedly appear in the search results. For example, a high number of papers referencing "ecosystems" were returned. While these papers might have included some of our search terms (i.e. "protected"), the scope of the REA was not to include natural resource valuation.

124 For example, in a sub-set of papers (303), 71% were below the relevancy threshold of 0.5 (i.e. using the search terms less frequently).

7.2.4 Quality Grading Criteria

Papers which met the criteria screened by the code and researcher investigation, were then screened by researchers for more complex criteria by grading the quality of the research papers based on the Quality Grading Criteria. All valid papers are included in the Evidence bank, regardless of Quality Criteria RAG rating. Although, we have included studies with low RAG ratings in the evidence bank, we provide strong caveats against following these methods and using values from these studies with low RAG rating.

Please note, that some studies in the literature reported more than one value. For example, some studies reported different values based using different methods for the same asset, other studies reported values for multiple assets. The Quality Grading Criteria was used to evaluate the methodological quality of each value, rather than the study as a whole. Thereby, values from the same study may have had differing grades.

This Quality Criteria RAG rating was outlined prior to the searches being conducted with specific exclusion criteria based on economic valuation methods. This criterion was written up in collaboration with stakeholders to determine rigorous standards for each evaluation method. A full set of criteria for low, medium, or high-quality rating is provided for each method in the evidence bank tab "2. Quality Grading Criteria QGC". In sum:

QGC1: Empirical design: Assessment of the survey design elements (in the case of CV) based on best practice from the literature[125] or the complexity of econometric modelling specification (in the case of DCEs, RP, or WV studies).[126]

QGC2: Method/Dataset: Assessment of the dataset on which the valuation is based, either in terms of the mechanisms by which the value is elicited (for instance, elicitation methods that encourage consequentiality and minimise hypothetical bias, in the case of SP), zonal or choice modelling (in the case of Travel Cost) or the size and reliability of the dataset on which econometric analysis is performed (in the case of Hedonic and WV). Furthermore, appropriate sample selection methods were taken into consideration, dependent upon the method used, to ensure values could be extrapolated to the wider population.

QGC3: Sample size (full sample): Assessment of the reliability of the study based on the size of the sample on which valuation is based. This will differ by non-market valuation method, due to the complexity of the econometric

[125] G Atkinson et al. 2018; Bateman et al. 2002; R. J. Johnston et al. 2017
[126] G Atkinson et al. 2018

modelling and power required, which can only be provided with higher sample sizes. Note that these thresholds should be interpreted as a loose 'rule of thumb' given that minimum sample size will affect the representativeness of the results and the confidence with which the researcher can extrapolate their findings to the wider population, and that will vary for each study depending on the relevant sample frame of individuals who are potentially affected.[127] In this study we grade minimum sample size based on overall sample available for each relevant user/non-user group.

The Quality Grading Criteria developed for this REA has never been attempted before in the cultural or environmental sector, and its synthesis of best practice guidance[128] and direct researcher experience represents a considerable contribution to the sector.

A final set of exclusions were applied by the researchers reviewing the final papers, based on whether the paper produced a valuation estimate that is commensurable with HM Treasury Green Book methods for non-market valuation.

The value of a good or service relates to the impact that it has on human welfare[129] and this can be expressed in terms of compensating or equivalent welfare measures as first devised by Hicks 1934.[130]

- **Compensating surplus (CS)** is the amount of money, paid or received, that will leave the agent in their initial welfare position following a change in the good.
- **Equivalent surplus (ES)** is the amount of money, to be paid or received, that will leave the agent in their subsequent welfare position in absence of a change in the good.

In essence, CS and ES refer to the change in income that holds welfare constant in light of the change in the provision of the good (which could be a change in the quantity and/or quality of the cultural heritage asset). We therefore exclude:

- Studies which employ SP methods but do not calculate a willingness-to-pay (WTP) value for the asset valued were excluded.
- Those studies using DCEs that do not calculate a WTP value for the whole asset, where it only values individual attributes or levels were excluded, even if these levels contained prices and were significant.

[127] For further guidance see Pearce and O'zdemiroglu 2002
[128] G Atkinson et al. 2018; Bateman et al. 2002; R. J. Johnston et al. 2017
[129] Freeman 2003
[130] Bockstael and McConnell 1980

- Studies using RP methods which do not calculate WTP values, or equivalent, for an asset but calculate demand function or price elasticities were excluded. Where studies provide only aggregate values (i.e., not individual or household level values), are listed in the evidence bank and are inflated to GBP£2020 prices, but do not equivalise values given difficulties of comparability of aggregate values.

Papers that meet all standards are therefore included in the evidence bank for the research synthesis (i.e., the written report).[131] However, we recommend that only those which meet the criteria threshold – i.e., those that were graded with medium to high research robustness using the Quality Grading Criteria – be taken into consideration for the research synthesis (i.e., the written report).[132]

132 We acknowledge that some of the grading criteria, such as those requiring more advanced modelling techniques, may necessarily exclude older papers from the list. We consider that this is appropriate, given that part of the purpose of this REA is to identify robust valuation estimates that could potentially be transferred to policy contexts with low risk of error.

Table 6-1 Summary of non-market valuation methods included in REA [133]

Stated Preference: Contingent Valuation	Stated Preference: Discrete Choice	Revealed Preference: Travel Cost	Revealed Preference: Hedonic (House) Prices	Benefit Transfer	Wellbeing Valuation
• Direct use value for people who actually visit/use/experience the cultural site/institution. • Indirect use value such as civic pride and social interaction. • Option value for people that may plan or want to use the cultural site/institution in the future. • Non-use value, including existence, altruistic, and bequest value.	• Direct use value for people who actually visit/use/experience the cultural site/institution. • Indirect use value such as civic pride and social interaction. • Option value for people that may plan or want to use the cultural site/institution in the future. • Non-use value, including existence, altruistic, and bequest value.	• TCM estimates direct user benefits. • It is unable to measure distant use value, indirect use value, option value, and non-use value. • Values are usually estimated as per visit values. These can be added up over the number of visitors per year to get an overall annual direct use value for the site/institution.	• Direct use value for people who actually visit the cultural site/institution is reflected in greater house prices closer to the asset. • Indirect use value, for example, the spillover effects from the presence of the cultural institution generates local benefits such as reduced crime and community cohesion. • Option value for people that are willing to pay to live near the cultural site/institution for the option to visit it if they so wish. • This method does not pick up non-use value as indirect use values are predominantly related to the	• BT is the method of transferring values from one study or site to another. Values will be defined by the original study, and can include the full range of use and non-use values as defined in the SP columns.	• Direct use value for people who visit a cultural site/institution (in a model that looks at visits). • Indirect use value such as reduced crime and community cohesion (in a model that looks at proximity to a cultural site/institution). • Option value for the option to visit a cultural site/institution if they wish (in a model that looks at proximity to a cultural site/institution). • This method does not usually pick up non-use value, although, there are ways in which a survey and study could be designed to pick up non-use value.

[133] informed by G Atkinson et al. 2018. Note that option value as defined in non-market valuation reflecting individual's decision-making about future use is distinct from the option value term is utilised elsewhere in the CBA literature, which relates to the benefits of waiting and learning about the social value of a resource.

Stated Preference: Contingent Valuation	Stated Preference: Discrete Choice	Revealed Preference: Travel Cost	Revealed Preference: Hedonic (House) Prices	Benefit Transfer	Wellbeing Valuation
			positive spillover benefits experienced by homeowners in the area surrounding the asset and not altruistic values towards the general public.		

7.2.5 Equivalisation of valuation results

Non-market values for culture and heritage assets can be elicited using a range of methods (SP, RP, WV) and the specifics of study design, survey questions, and modelling specification for each method will dictate how the reported value should be interpreted. For instance, an SP survey can elicit willingness to pay for a cultural asset as a one-off, monthly or annual payment, on behalf of the individual or household, or as an aggregate for the population of interest. Furthermore, valuation studies from different countries will be based in those currencies and are subject to different budget constraints from differences in costs of living. These issues should be taken into account when converting to GBP. To ensure that the values obtained from different time periods and jurisdictions are fully comparable across the evidence bank, we performed the following adjustments.[134]

1. Adjust for purchasing power parity by eliminating differences in price levels between countries. This was achieved by using purchasing power parities (PPPs) available from the OECD[135]. Using these PPPs we were able to get all valuations into GBP, adjusted for purchasing power, for the year of the study.
2. The next step was to adjust for inflation and bring all of the values up to GBP 2020. To do this we used UK CPI indices available from ONS[136].

Some studies did not specify the year that their study took place. In this instance, we assumed a study date of one year before the paper was published and uprated from there. Accounting for differences in the cost of living between countries was a crucial step because respondents in countries with a higher cost of living will have more constrained household budgets and may be less likely to report a higher WTP value compared to respondents in countries with a lower cost of living because they have less purchasing power. All of which affects the marginal utility of income and a person's ability to pay.

To equivalise units (i.e., whether the value was per person or per household) and allow comparison between studies, we converted all valuations to individual-level. This was done by dividing per-household valuations by the average number of people per household in the country the study was conducted on. This data was primarily sourced from Eurostat[137]. Some papers did not explicitly state whether their WTP estimate was per individual or per

134 The functions that underlie each of these adjustments are contained within the evidence bank Excel spreadsheet.

135 PPPs were taken from https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm

136 CPI indices were taken from https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7bt/mm23

137 Data for countries not in Europe was sourced from the Australian Institute of Family Studies (Australia), the UN (South Africa) and the US Census Bureau (US). Moreover, data for England and Scotland separately was sourced from the ONS and the Scottish Government, respectively.

household. In this instance, we assumed the valuation was per individual since this is more common in the literature.[138]

It is worth noting that some assumptions were needed in order to arrive at comparable valuations:

- Some papers gave a range as their WTP valuation rather than a single value. In this case, we used the midpoint of the range.
- When different models were used to estimate the WTP value for the same asset, we used the valuation from the model the author deemed the most robust (through explicit endorsement or implicit endorsement by discussing one model over other model/s). If no indication of robustness was given, we take a midpoint estimate of all reported model values.
- When different methodologies were used, we used the Quality Grading Criteria guidance to access which was the most robust.
- Often, both an average and a median WTP values were given. When this was the case, we reported both values but only conducted analysis on the average value. Using the mean WTP value rather than the median is good practice in CV studies Darling and et al 2000. The mean is relevant if the context of the valuation exercise is cost benefit analysis because it represents an average WTP for the population which can be aggregated (by the population size) to derive the total WTP across the population. Unless otherwise stated, we report mean WTP throughout this paper.
- When two average valuations were given in a paper that measured the WTP for the same asset but a different subset of respondents (e.g., low income and high income), we reported both values and conducted analysis in the midpoint of the valuations.

138 Note that unit equivalisation was not performed for the handful of studies which provided only aggregate values, because it was not clear in all cases what the appropriate population for de-aggregation should be.

7.3 Full tables

7.3.1 Stated Preference: Contingent Valuation

Table 6-2 Culture and heritage values 2000-2019 using Stated Preference Contingent Valuation methods, by asset type

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Archaeological asset	5	5	1	1	6	6
Archaeological asset	3	3	1	1	4	4
Ruin	2	2	0	0	2	2
Art engagement	0	0	1	0	1	0
Art engagement	0	0	1	0	1	0
Built heritage	3	3	6	4	9	7
Built heritage	0	0	4	4	4	4
Castle	2	2	1	0	3	2
Cinema	0	0	0	0	0	0
City	0	0	0	0	0	0
High street	0	0	0	0	0	0
Stately home	1	1	1	0	2	1
Town hall	0	0	0	0	0	0
Village	0	0	0	0	0	0
Cultural institution	37	23	25	13	62	36
Archive	6	1	3	0	9	1
Concert hall	2	0	0	0	2	0
Festival	4	4	2	2	6	6
Gallery	2	1	8	4	10	5
Library	10	10	3	1	13	11
Museum	13	7	9	6	22	13
Music venue	0	0	0	0	0	0
Theatre	0	0	0	0	0	0
Digital asset	2	1	2	2	4	3
Digital asset	0	0	0	0	0	0
Public service	2	1	2	2	4	3

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
broadcasting						
Industrial heritage	0	0	0	0	0	0
Aqueduct	0	0	0	0	0	0
Bridge	0	0	0	0	0	0
Canal	0	0	0	0	0	0
Dock	0	0	0	0	0	0
Factory	0	0	0	0	0	0
Harbour	0	0	0	0	0	0
Mine	0	0	0	0	0	0
Port	0	0	0	0	0	0
Quarry	0	0	0	0	0	0
Rail	0	0	0	0	0	0
Warehouse	0	0	0	0	0	0
Watermill	0	0	0	0	0	0
Waterway	0	0	0	0	0	0
Windmill	0	0	0	0	0	0
Historical amenities	3	3	2	2	5	5
Garden	1	1	2	2	3	3
Historical amenities	2	2	0	0	2	2
Monument	0	0	0	0	0	0
Plaque	0	0	0	0	0	0
Public art	0	0	0	0	0	0
Protected area	1	1	0	0	1	1
Protected area	1	1	0	0	1	1
Protected business	0	0	0	0	0	0
Protected city	0	0	0	0	0	0
Protected structure	0	0	0	0	0	0
Protected town	0	0	0	0	0	0
Protected transport	0	0	0	0	0	0
Religious asset	1	1	0	0	1	1
Religious asset	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Cathedral	1	1	0	0	1	1
Total	52	37	37	22	89	59

Table 6-3 Contingent valuation methodology: Elicitation method (per WTP value)

Elicitation method	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Open-end	10	1	18	3	28	4
Payment ladder	14	8	15	15	29	23
Single-bounded dichotomous choice	9	9	1	1	10	10
Double-bounded dichotomous choice	19	19	1	1	20	20
Dichotomous choice (single/double unspecified)	0	0	2	2	2	2

Table 6-4 Contingent valuation methodology: Payment vehicle (per WTP value)

Payment vehicle	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Not specified	1	1	1	0	2	1
Donation	11	11	19	7	30	18
Entry/membership fee	11	5	1	1	12	6
Obligatory one-time payment	2	2	0	0	2	2
Subscription	2	1	0	0	2	1
License fee	0	0	2	2	2	2
Tax	23	15	9	9	32	24
Compensation (WTA)	2	2	5	3	7	5

Table 6-5 Contingent valuation methodology: Payment term (per WTP value)

Payment term	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Elicitation method	Total (any quality)	Good quality (Medium-high quality rating)
One-off	9	9	13	9	22	18
Per visit	8	2	2	2	10	4
Per season	2	2	0	0	2	2
Fixed term	2	2	2	2	4	4
Indefinite recurring	22	14	18	9	40	23
Not specified	9	8	2	0	11	8

Table 6-6 Contingent valuation methodology: Aggregation factors: individual or household)

Aggregation factors	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Elicitation method	Total (any quality)	Good quality (Medium-high quality rating)
Household	5	4	3	3	8	7
Individual	17	10	27	17	44	27
Individual or business	0	0	2	2	2	2
Not specified	30	23	5	0	35	23

Reliable value estimates require minimum sample size requirements to be met. Naturally, there is a trade-off between research cost and accuracy, as the entire target population sample cannot be surveyed due to access issues and cost, among other things. While smaller sample sizes are permissible if more information has been gained from each survey respondent, samples under 200 respondents should not be considered without careful consideration. However, if the survey requires more sub-samples, split over different survey versions for example, then a greater sample size is required (a suggested 250-500 is required for each sub-sample in an open-ended survey and 500-1000 for a closed-ended survey; Pearce et al. 2002). As a general rule of thumb, sample sizes of less than 200 respondents are not considered to have robust findings in CV methods as this does not account for variance in the underlying target population. Thereby, any values with sample sizes of at least 200 respondents provide an acceptable estimate of a target population and received a medium-high quality grading on our criteria (QGC3).

Most values obtained in the REA were of adequate sample size (64 in total, with 52 of these values being of medium-high quality). 25 values reported inadequate sampling (<200). While 7 of these values received medium-high quality ratings, their poorer sample size was compensated with higher quality CV study designs (Appendix Table 6-7).

Table 6-7 Contingent valuation methodology: Sample size (per WTP value)

Sample size	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Elicitation method	Total (any quality)	Good quality (Medium-high quality rating)
<200	15	7	10	0	25	7
200-500	15	13	7	4	22	17
500+	22	17	20	18	42	35
Not specified	0	0	0	0	0	0

7.3.2 Stated Preference: Discrete Choice Experiments

Table 6-8 Culture and heritage values 2000-2019 using Stated Preference Discrete Choice Method, by asset type

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Archaeological asset	11	10	0	0	11	10
Archaeological asset	5	4	0	0	5	4
Ruin	6	6	0	0	6	6
Art engagement	0	0	0	0	0	0
Art engagement	0	0	0	0	0	0
Built heritage	3	3	0	0	3	3
Built heritage	3	3	0	0	3	3
Castle	0	0	0	0	0	0
Cinema	0	0	0	0	0	0
City	0	0	0	0	0	0
High street	0	0	0	0	0	0
Stately home	0	0	0	0	0	0
Town hall	0	0	0	0	0	0
Village	0	0	0	0	0	0
Cultural institution	10	10	0	0	10	10
Archive	0	0	0	0	0	0
Concert hall	0	0	0	0	0	0
Festival	0	0	0	0	0	0
Gallery	0	0	0	0	0	0
Library	0	0	0	0	0	0
Museum	9	9	0	0	9	9
Music venue	0	0	0	0	0	0
Theatre	1	1	0	0	1	1
Digital asset	0	0	4	2	4	2
Digital asset	0	0	0	0	0	0
Public service	0	0	4	2	4	2

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
broadcasting						
Industrial heritage	0	0	0	0	0	0
Aqueduct	0	0	0	0	0	0
Bridge	0	0	0	0	0	0
Canal	0	0	0	0	0	0
Dock	0	0	0	0	0	0
Factory	0	0	0	0	0	0
Harbour	0	0	0	0	0	0
Mine	0	0	0	0	0	0
Port	0	0	0	0	0	0
Quarry	0	0	0	0	0	0
Rail	0	0	0	0	0	0
Warehouse	0	0	0	0	0	0
Watermill	0	0	0	0	0	0
Waterway	0	0	0	0	0	0
Windmill	0	0	0	0	0	0
Historical amenities	5	2	0	0	5	2
Garden	0	0	0	0	0	0
Historical amenities	1	1	0	0	1	1
Monument	4	1	0	0	4	1
Plaque	0	0	0	0	0	0
Public art	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected business	0	0	0	0	0	0
Protected city	0	0	0	0	0	0
Protected structure	0	0	0	0	0	0
Protected town	0	0	0	0	0	0
Protected transport	0	0	0	0	0	0
Religious asset	0	0	0	0	0	0
Religious asset	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Cathedral	0	0	0	0	0	0
Total	29	25	4	2	33	27

Table 6-9 Discrete Choice methodology: DCE Design

DCE Design	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Simple model	0	0	0	0	0	0
Under 8 attributes with less than 4 levels each	29	25	0	0	29	25
More than 8 attributes and/or more than 4 levels	0	0	4	2	4	2

Within the academic literature, most values were obtained using advanced regression modelling (24), with all but one being rated high-quality by the Quality Criteria Rating. Within the academic literature only 5 values use a binary logit model (with 3 of these graded low-quality by the Quality Criteria Rating).

In the grey literature, 2 of the 4 DCE values were obtained using advanced multinomial logit models (n=2), all rated medium-high quality by the Quality Criteria Rating. The remaining 2 values used the simple binary logit model and were accordingly rated low quality by the Quality Criteria Rating.

Table 6-10 Discrete Choice methodology: Regression modelling

Regression Modelling	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Binary logit	5	2	2	0	7	2
Conditional logit	6	6	0	0	6	6
Latent class	3	3	0	0	3	3

Regression Modelling	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Mixed logit	9	8	0	0	9	8
Nested logit	6	6	0	0	6	6
Multinomial logit model	0	0	2	2	2	2

All values with orthogonal (12), d-optimal (3) and other efficient designs (2) were rated highly in the Quality Criteria Rating due to high-quality Discrete Choice design and regression modelling.

A number of studies did not provide details on the design underlying their values (17), although 11 of these values were rated as medium-high quality due to other factors in their design being of higher quality. A higher proportion of DCE values from the grey literature did not specify the design efficiencies of their DCE design (4 out of 4). However, nearly half of the values from the academic literature (13 of 29) failed to specify the efficiency of their DCE design. This is important information required to understand how welfare consistent the DCE values are.

Table 6-11 Discrete Choice methodology: Welfare consistent design

Welfare Consistent Design	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Not specified	13	9	4	2	17	11
Orthogonal design	12	12	0	0	12	12
D optimal design	3	3	0	0	3	3
Efficient design (other)	1	1	0	0	1	1

In the REA, only 3 values included status quo and opt-out options (all rated medium-high quality). However, it is not necessary to include both elements in DCE design for values to be considered consistent with welfare theory. 11 values included only an opt out choice (10 rated medium-high quality), and 13 included only status quo (11 rated medium-high quality).

A small number (n=5) of academic peer-reviewed values applied simple Paired comparison design which does not allow for an opt out option. Only two of

these were considered methodologically robust on enough design considerations to be rated medium-high quality by the Quality Criteria Rating.

In the grey literature, 2 of the 4 values applied simple Paired comparison, with the remaining three applying a Status quo option only. However, based on other quality criteria, all 2 of the paired comparison values were rated medium-high quality, by the Quality Criteria Rating, while none of those values containing a Status quo option was rated medium-high quality by the Quality Criteria Rating.

Table 6-12 Discrete Choice methodology: Welfare consistency

Zonal/individual data	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Paired comparison (i.e., no opt out option)	5	2	2	2	7	4
Opt-out	11	10	0	0	11	10
Status quo	10	10	2	0	12	10
Status quo and opt-out	3	3	0	0	3	3

Greater sample sizes (e.g., =>500) are required under Discrete Choice models, compared to CV methods, because of the number of attributes processed in the regression modelling. Smaller sampling under Discrete Choice designs presents issues, as these samples are overly reliant on respondents to provide more information (i.e., provide their preferences for each of the attributes presented). As such, smaller sampling does not account for inattentiveness and fatigue effects, which would lead to inaccurate estimates. In such scenarios, welfare consistent designs can help to automatically group attributes to maximise choice efficiency.

Just under half (n=15) of DCE values for culture and heritage assets were based on sample sizes greater than 500, all of which were rated medium-high quality by the Quality Criteria Rating (Appendix Table 6-13).

Over half of all DCE values (18) recorded sample sizes below 500. Of these, 12 were still graded medium-high quality overall, based on their overarching empirical design. We note that sample size thresholds for DCE in the REA should be treated as a rule of thumb, and full assessment of the adequacy of sample size for calculation of WTP in DCE depends on a number of factors, including sample frame, number of attributes and number of levels, which should be assessed in detail on a case by case basis.

Table 6-13 DCE methodology: Sample size (per WTP value)

Sample size	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
<500	16	12	2	0	18	12
500-2,000	13	13	2	2	15	15
2,000+	0	0	0	0	0	0

7.3.3 Revealed Preference: Travel Cost Methods

Table 6-14 Culture and heritage values 2000-2019 using Revealed Preference Travel Cost Method, by asset type

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Archaeological asset	1	1	0	0	1	1
Archaeological asset	1	1	0	0	1	1
Ruin	0	0	0	0	0	0
Art engagement	0	0	0	0	0	0
Art engagement	0	0	0	0	0	0
Built heritage	2	0	0	0	2	0
Built heritage	0	0	0	0	0	0
Castle	0	0	0	0	0	0
Cinema	0	0	0	0	0	0
City	1	0	0	0	1	0
High street	0	0	0	0	0	0
Stately home	0	0	0	0	0	0
Town hall	0	0	0	0	0	0
Village	1	0	0	0	1	0
Cultural institution	7	2	2	0	9	2
Archive	0	0	0	0	0	0
Concert hall	1	0	0	0	1	0
Festival	1	0	0	0	1	0
Gallery	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Library	0	0	0	0	0	0
Museum	5	2	2	0	7	2
Music venue	0	0	0	0	0	0
Theatre	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Public service broadcasting	0	0	0	0	0	0
Industrial heritage	0	0	0	0	0	0
Aqueduct	0	0	0	0	0	0
Bridge	0	0	0	0	0	0
Canal	0	0	0	0	0	0
Dock	0	0	0	0	0	0
Factory	0	0	0	0	0	0
Harbour	0	0	0	0	0	0
Mine	0	0	0	0	0	0
Port	0	0	0	0	0	0
Quarry	0	0	0	0	0	0
Rail	0	0	0	0	0	0
Warehouse	0	0	0	0	0	0
Watermill	0	0	0	0	0	0
Waterway	0	0	0	0	0	0
Windmill	0	0	0	0	0	0
Historical amenities	2	2	0	0	2	2
Garden	1	1	0	0	1	1
Historical amenities	1	1	0	0	1	1
Monument	0	0	0	0	0	0
Plaque	0	0	0	0	0	0
Public art	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected business	0	0	0	0	0	0
Protected city	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Protected structure	0	0	0	0	0	0
Protected town	0	0	0	0	0	0
Protected transport	0	0	0	0	0	0
Religious asset	1	0	0	0	1	0
Religious asset	0	0	0	0	0	0
Cathedral	1	0	0	0	1	0
Total	13	5	2	0	15	5

Table 6-15 Travel Cost Methodology Single/Multi-purpose travel (per value)

Single/Multi-purpose	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Single-purpose travel	5	1	2	0	7	1
Single-purpose travel (sample only includes those traveling for that purpose)	4	1	0	0	4	1
Single-purpose travel (assumes sample only traveling for that purpose)	2	1	0	0	2	1
Single-purpose travel (with assumptions for dealing with multi-trips)	0	0	0	0	0	0
Multi-purpose travel	2	2	0	0	2	2

Table 6-16 Travel Cost Methodology Model Covariates (per value)

Single/Multi-purpose	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
No covariates	3	0	2	0	5	0
Simple covariates	3	0	0	0	3	0
Standard covariates	7	5	0	0	7	5

Of 13 academic peer-reviewed values, only 5 values met the high-quality standard in the academic literature. High quality values using primary (individual) TCM data (4) accounted for heterogeneity through mixed logit models, employed GIS data for more reliable travel data, included dummy variables for multi-purpose visits, whilst logging card holder visits. The values obtained from zonal datasets were of low-quality due to the lack of primary data collection. However, one zonal value was rated as high-quality as it attempted to control for multi-purpose visits. In the grey literature, both TCM values were based on zonal data and rated low quality.

Table 6-17 Travel Cost Methodology Zonal/Individual data (per value)

Zonal/Individual data	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Zonal	9	1	2	0	11	1
Individual TCM using primary data	4	4	0	0	4	4

Around a third of academic peer-reviewed TCM studies are based on a sample size below the minimum recommended threshold of 250, while the same number are only at the medium quality threshold for sample size. This may reflect the fact that the majority of studies rely on low quality zonal data, which may be subject to sample size restrictions (Appendix Table 6-18).

In the grey literature, one study did not specify its sample size, which is a major problem for analysts looking to assess the reliability and quality of culture and heritage values obtained through TCM.[139]

Nine values in the academic literature obtained a minimum sample of 250 or more, with only 4 of these values of medium-high quality. One notable value was obtained using automated logging of individual visitor data, allowing for a very large sample size (69,643), which was accompanied with GIS data to present a robust model. Values with adequate sampling but poor methodological quality, and the four values which obtained an inadequate sample size, were marked down. These included the employment of zonal data rather than individual data, or the lack of consideration into multi-purpose visits.

In the grey literature one TCM study was based on a large sample of 1,000+ and one study did not specify its sample size, which is a major problem for analysts looking to assess the reliability and quality of culture and heritage values obtained through TCM.

Table 6-18 Travel Cost Methodology: Sample size (per value)

Sample size	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
<250	4	1	0	0	4	1
250-999	5	1	0	0	5	1
1,000+	4	3	1	0	5	3
Not specified	0	0	1	0	1	0

[139] We note that the minimum sample sizes recommended in the Quality Criteria Rating were intended as rules of thumb. Sample size need to be calculated on a case by case basis, factoring in target sample groups and maximum sample frame considerations. We would recommend deeper review of the low-quality rated studies to ascertain whether appropriate sample size calculations were undertaken, and in such cases the quality rating could be increased to medium or high. While there is often a trade-off in precision of estimates and cost of sampling, the required size of sampling may be partially dependent on the specific travel cost approach used. For example, zonal data provides simpler data collection as it can be obtained from card holder logs, compared to individual data collection which would require active field research. At the same time, the reliance on estimation in zonal data sets and the concerns of inaccurate recall in individual data require greater sampling for more accurate estimates. A suggested minimum sample of 250 individuals is accepted for travel costs methods to have be graded medium-quality.

7.3.4 Revealed Preference: Hedonic Pricing

Table 6-19 Culture and heritage values 2000-2019 using Revealed Preference Hedonic Pricing Method, by asset type

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Archaeological asset	0	0	0	0	0	0
Archaeological asset	0	0	0	0	0	0
Ruin	0	0	0	0	0	0
Art engagement	0	0	0	0	0	0
Art engagement	0	0	0	0	0	0
Built heritage	1	1	0	0	1	1
Built heritage	1	1	0	0	1	1
Castle	0	0	0	0	0	0
Cinema	0	0	0	0	0	0
City	0	0	0	0	0	0
High street	0	0	0	0	0	0
Stately home	0	0	0	0	0	0
Town hall	0	0	0	0	0	0
Village	0	0	0	0	0	0
Cultural institution	0	0	0	0	0	0
Archive	0	0	0	0	0	0
Concert hall	0	0	0	0	0	0
Festival	0	0	0	0	0	0
Gallery	0	0	0	0	0	0
Library	0	0	0	0	0	0
Museum	0	0	0	0	0	0
Music venue	0	0	0	0	0	0
Theatre	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Public service broadcasting	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Industrial heritage	0	0	1	1	1	1
Aqueduct	0	0	0	0	0	0
Bridge	0	0	0	0	0	0
Canal	0	0	1	1	1	1
Dock	0	0	0	0	0	0
Factory	0	0	0	0	0	0
Harbour	0	0	0	0	0	0
Mine	0	0	0	0	0	0
Port	0	0	0	0	0	0
Quarry	0	0	0	0	0	0
Rail	0	0	0	0	0	0
Warehouse	0	0	0	0	0	0
Watermill	0	0	0	0	0	0
Waterway	0	0	0	0	0	0
Windmill	0	0	0	0	0	0
Historical amenities	3	3	0	0	3	3
Garden	0	0	0	0	0	0
Historical amenities	3	3	0	0	3	3
Monument	0	0	0	0	0	0
Plaque	0	0	0	0	0	0
Public art	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected business	0	0	0	0	0	0
Protected city	0	0	0	0	0	0
Protected structure	0	0	0	0	0	0
Protected town	0	0	0	0	0	0
Protected transport	0	0	0	0	0	0
Religious asset	0	0	0	0	0	0
Religious asset	0	0	0	0	0	0
Cathedral	0	0	0	0	0	0
Total	4	4	1	1	5	5

Most values (4 of 5) accounted for both structural and spatial covariates. For the value that only included structural covariates, the regression models included basic housing characteristics, but only some models included other variables, such as garage spaces, which may have biased coefficients if inconsistently collected and put into the model.

Table 6-20 Revealed Preference methodology: Hedonic Pricing Model covariates

Model covariates	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Structural covariates only	1	1	0	0	1	1
Structural and spatial covariates	3	3	1	1	4	4

Table 6-21 Revealed Preference methodology: Hedonic Pricing Regression Model

Regression Model	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
OLS	1	1	0	0	1	1
Log linear OLS	3	3	0	0	3	3
OLS plus Difference in Difference	0	0	1	1	1	1

Table 6-22 Revealed Preference methodology: Hedonic Pricing Data Source

Data Source	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Administrative data	3	3	0	0	3	3
Real estate data	0	0	0	0	0	0
Real estate data enriched with administrative data	1	1	1	1	2	2

Large datasets (>=1000) are required to be confident in the values elicited using HP methods. Fortunately, large datasets are widely available from real estate and administrative resources online. As such, all values in the REA used large administrative data sets (with one value enhancing the dataset with real estate data) therefore all received high-quality grading.

All hedonic values had a high quality rating around the size of the data samples used. This suggests that the datasets available for hedonic analysis are strong across multiple jurisdictions, and we would expect this to continue into the future with the opening up of online land registry and house market data. This has resulted in a healthy set of hedonic values for those cultural heritage asset types which are amenable to this method (i.e., those assets which are expected to affect local house prices).

Table 6-23 Hedonic Pricing methodology: Sample size (per WTP value)

Sample size	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
<250	0	0	0	0	0	0
250-999	0	0	0	0	0	0
1,000+	4	4	1	1	5	5

7.3.5 Benefit Transfer

Table 6-24 Culture and heritage values 2000-2019 using Benefit Transfer, by asset type

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Archaeological asset	1	0	0	0	1	0
Archaeological asset	1	0	0	0	1	0
Ruin	0	0	0	0	0	0
Art engagement	0	0	0	0	0	0
Art engagement	0	0	0	0	0	0
Built heritage	0	0	9	2	9	2

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Built heritage	0	0	3	2	3	2
Castle	0	0	6	0	6	0
Cinema	0	0	0	0	0	0
City	0	0	0	0	0	0
High street	0	0	0	0	0	0
Stately home	0	0	0	0	0	0
Town hall	0	0	0	0	0	0
Village	0	0	0	0	0	0
Cultural institution	0	0	2	2	2	2
Archive	0	0	0	0	0	0
Concert hall	0	0	0	0	0	0
Festival	0	0	0	0	0	0
Gallery	0	0	0	0	0	0
Library	0	0	0	0	0	0
Museum	0	0	2	2	2	2
Music venue	0	0	0	0	0	0
Theatre	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Public service broadcasting	0	0	0	0	0	0
Industrial heritage	0	0	0	0	0	0
Aqueduct	0	0	0	0	0	0
Bridge	0	0	0	0	0	0
Canal	0	0	0	0	0	0
Dock	0	0	0	0	0	0
Factory	0	0	0	0	0	0
Harbour	0	0	0	0	0	0
Mine	0	0	0	0	0	0
Port	0	0	0	0	0	0
Quarry	0	0	0	0	0	0
Rail	0	0	0	0	0	0
Warehouse	0	0	0	0	0	0
Watermill	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Waterway	0	0	0	0	0	0
Windmill	0	0	0	0	0	0
Historical amenities	0	0	0	0	0	0
Garden	0	0	0	0	0	0
Historical amenities	0	0	0	0	0	0
Monument	0	0	0	0	0	0
Plaque	0	0	0	0	0	0
Public art	0	0	0	0	0	0
Protected area	0	0	2	2	2	2
Protected area	0	0	0	0	0	0
Protected business	0	0	0	0	0	0
Protected city	0	0	2	2	2	2
Protected structure	0	0	0	0	0	0
Protected town	0	0	0	0	0	0
Protected transport	0	0	0	0	0	0
Religious asset	0	0	2	2	2	2
Religious asset	0	0	0	0	0	0
Cathedral	0	0	2	2	2	2
Total	1	0	15	8	16	8

Table 6-25 Benefit Transfer methodology: Data collection

Data collection	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Meta-review	1	0	0	0	1	0
Primary data collection	0	0	15	8	15	8

Table 6-26 Benefit Transfer methodology: Study sites

Study sites	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
1 study site	0	0	1	0	1	0
2 study sites	0	0	6	0	6	0
3 study sites	1	0	1	1	2	1
4 or more study sites	0	0	7	7	7	7

Table 6-27 Benefit Transfer methodology: Transfer testing

Transfer testing	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
No transfer testing	1	0	1	0	2	0
Full suite of Unit, Adjusted, Function testing	0	0	14	8	14	8

7.3.6 Wellbeing Valuation

Table 6-28 Culture and heritage values 2000-2019 using Wellbeing Valuation, by asset type

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Archaeological asset	0	0	0	0	0	0
Archaeological asset	0	0	0	0	0	0
Ruin	0	0	0	0	0	0
Art engagement	1	0	3	3	4	3
Art engagement	1	0	3	3	4	3

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Built heritage	0	0	0	0	0	0
Built heritage	0	0	0	0	0	0
Castle	0	0	0	0	0	0
Cinema	0	0	0	0	0	0
City	0	0	0	0	0	0
High street	0	0	0	0	0	0
Stately home	0	0	0	0	0	0
Town hall	0	0	0	0	0	0
Village	0	0	0	0	0	0
Cultural institution	0	0	4	4	4	4
Archive	0	0	0	0	0	0
Concert hall	0	0	0	0	0	0
Festival	0	0	0	0	0	0
Gallery	0	0	0	0	0	0
Library	0	0	1	1	1	1
Museum	0	0	3	3	3	3
Music venue	0	0	0	0	0	0
Theatre	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Digital asset	0	0	0	0	0	0
Public service broadcasting	0	0	0	0	0	0
Industrial heritage	0	0	0	0	0	0
Aqueduct	0	0	0	0	0	0
Bridge	0	0	0	0	0	0
Canal	0	0	0	0	0	0
Dock	0	0	0	0	0	0
Factory	0	0	0	0	0	0
Harbour	0	0	0	0	0	0
Mine	0	0	0	0	0	0
Port	0	0	0	0	0	0
Quarry	0	0	0	0	0	0
Rail	0	0	0	0	0	0

Asset	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Warehouse	0	0	0	0	0	0
Watermill	0	0	0	0	0	0
Waterway	0	0	0	0	0	0
Windmill	0	0	0	0	0	0
Historical amenities	0	0	5	5	5	5
Garden	0	0	0	0	0	0
Historical amenities	0	0	5	5	5	5
Monument	0	0	0	0	0	0
Plaque	0	0	0	0	0	0
Public art	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected area	0	0	0	0	0	0
Protected business	0	0	0	0	0	0
Protected city	0	0	0	0	0	0
Protected structure	0	0	0	0	0	0
Protected town	0	0	0	0	0	0
Protected transport	0	0	0	0	0	0
Religious asset	0	0	0	0	0	0
Religious asset	0	0	0	0	0	0
Cathedral	0	0	0	0	0	0
Total	1	0	12	12	13	12

Table 6-29 Wellbeing Valuation methodology: Empirical design

Site visit	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Income instrumented for	0	0	9	9	9	9
Fixed Effects	0	0	0	0	0	0

Site visit	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Both approaches utilised	0	0	0	0	0	0
Endogeneity of income not adequately addressed	1	0	3	3	4	3

Table 6-30 Wellbeing Valuation methodology: Sample group

Site visit	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
Regular engagement with good being valued	0	0	3	3	0	0
Possible infrequent engagement with good being valued	1	0	9	9	1	0

Compared to the previous valuation methods, WV requires very large sample sizes to be considered a medium- (1,000+) to high-quality (10,000+) study design. WV relies on the statistical estimation of the impact of a non-market good on life satisfaction and large sample sizes are required to uncover impacts such as these. As such, large national datasets are typically used as opposed to primary data. Indeed this is the case with all of the WV values graded in this study, bar one (the single academic value).

Given this, the academic value was graded as low-quality in terms of sample size. Otherwise we noted good sample sizes in the grey literature with ten graded as high quality and two as medium.

Table 6-31 Wellbeing valuation methodology: Sample size (per WTP value)

Sample size	Academic (peer-reviewed)		Grey literature (not all peer-reviewed)		All papers	
	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)	Total (any quality)	Good quality (Medium-high quality rating)
<500	1	0	0	0	1	0
1,000-10,000	0	0	2	2	2	2
10,000+	0	0	10	10	10	10

8 Bibliography

Alexandros, Apostolakis, and Shabbar Jaffry. 2005. 'Stated Preferences for Two Cretan Heritage Attractions'. *Annals of Tourism Research* 32 (4):985–1005. https://doi.org/10.1016/j.annals.2005.01.004.

Atkinson, G, N.A. Braathen, B. Groom, and S Mourato. 2018. 'Cost-Benefit Analysis and the Environment - Further Developments and Policy Use'. Berlin, Germany: Organisation for Economic Co-Operation and Development (OECD). http://www.oecd.org/environment/cost-benefit-analysis-and-the-environment-9789264085169-en.htm.

Atkinson, Giles, Ben Groom, Nicholas Hanley, and Susana Mourato. 2018. 'Environmental Valuation and Benefit-Cost Analysis in U.K. Policy'. *Journal of Benefit-Cost Analysis* 9 (1):97–119. https://doi.org/10.1017/bca.2018.6.

Australia Council for the Arts. 2015. 'Arts Nation: An Overview of Australian Arts'. Sydney, Australia. https://www.australiacouncil.gov.au/workspace/uploads/files/arts-nation-october-2015-5638269193891.pdf.

Bakhshi, H., Daniel Fujiwara, R. N. Lawton, Susana Mourato, and Paul Dolan. 2015. 'Measuring Economic Value in Cultural Institutions'. Cultural Value Project. London, UK: Arts and Humanities Research Council. https://ahrc.ukri.org/documents/project-reports-and-reviews/measuringeconomicvalue/.

Bateman, I., R. T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, et al. 2002. *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham, UK: Edward Elgar.

Beel, Jöran, and Bela Gipp. 2009. 'Google Scholar's Ranking Algorithm: An Introductory Overview'. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, 1:230–41. Rio de Janeiro (Brazil).

Belot, Michèle, and Sjef Ederveen. 2012. 'Cultural Barriers in Migration between OECD Countries'. *Journal of Population Economics* 25 (3):1077–1105.

Boardman, Anthony, David Greenberg, Aidan Vining, David Weimer, and & 1 more. 2010. *Cost-Benefit Analysis*. 4 edition. Boston, MA: Prentice Hall.

Bockstael, Nancy E., and Kenneth E. McConnell. 1980. 'Calculating Equivalent and Compensating Variation for Natural Resource Facilities'. *Land Economics*, 56–63.

Brouwer, R., and S. Navrud. 2015. 'The Use and Development of Benefit Transfer in Europe'. In *Benefit Transfer of Environmental and Resource Values - A Guide for Researchers and Practitioners*, edited by R. Johnston, J. Rolfe, R.S. Rosenberger, and R. Brouwer, 71–84. The Economics of Non-Market Goods and Resources 14. London, UK: Springer. http://www.springer.com/gb/book/9789401799294.

Brown, Thomas C, and Robin Gregory. 1999. 'Why the WTA-WTP Disparity Matters'. *Ecological Economics* 28 (3):323–35. https://doi.org/10.1016/S0921-8009(98)00050-0.

Chau, Kwong Wing, and T. L. Chin. 2003. 'A Critical Review of Literature on the Hedonic Price Model'. *International Journal for Housing Science and Its Applications* 27 (2):145–65.

Choi, Andy S., Brent W. Ritchie, Franco Papandrea, and Jeff Bennett. 2010. 'Economic Valuation of Cultural Heritage Sites: A Choice Modeling Approach'. *Tourism Management* 31 (2):213–20. https://doi.org/10.1016/j.tourman.2009.02.014.

Clark, Andrew, Sarah Flèche, Richard Layard, Nattavudh Powdthavee, and George Ward. 2018. *The Origins of Happiness: The Science of Well-Being Over the Life Course*. Princeton: Princeton University Press.

Corruchaga, J. L. A., and Pilar Fatás Monforte. 2006. 'The New Museum of the Altamira: Finding Solutions to Tourism Pressure'. *Of The Past For Future Integration Archaeology and Conservation: İçinde*, 177–83.

Crossick, G, and P Kaszynska. 2016. 'Understanding the Value of Arts & Culture. The AHRC Cultural Value Project'. London, UK: AHRC.

Darling, Arthur C., and et al. 2000. 'Uncertainty in Cost-Benefit Analysis Based on Referendum Contingent Valuation'. *Mpact Assessment and Project Appraisal*, 18 (2):125–37. https://doi.org/10.3152/147154600781767466.

DCMS. 2006. 'Willingness to Pay for the BBC during the next Charter Period'. London, UK: Department for Culture, Media and Sport.

Del Saz Salazar, S., and J. Montagud Marques. 2005. 'Valuing Cultural Heritage: The Social Benefits of Restoring and Old Arab Tower'. *Journal of Cultural Heritage* 6 (1):69–77. https://doi.org/10.1016/j.culher.2004.09.001.

Delaney, L, and F O'Toole. 2004. 'Irish Public Service Broadcasting: A Contingent Valuation Analysis'. *The Economic and Social Review* 35 (3):321–50.

Freeman, A.M. 2003. 'Economic Valuation: What and Why'. In *A Primer on Nonmarket Valuation*, by P.A. Champ, K.J. Boyle, and T.C. Brown, 1–25. Kluwer Academic Publishers.

Fujiwara, D., H. Bakhshi, S. Mourato, R. Lawton, U. Hotopp, and J. Davies. 2018. 'The Economic Value of Culture: A Benefit Transfer Study'. London, UK: Department for Digital Culture Media and Sport.

Fujiwara, D., T. Cornwall, and P. Dolan. 2014. 'Heritage and Wellbeing'. London, UK: SImetrica, English Heritage.

Fujiwara, D., L. Kudrna, and P. Dolan. 2014. 'Quantifying and Valuing the Wellbeing Impacts of Culture and Sport'. London, UK: Department for Culture, Media & Sport.

Fujiwara, D., P. Oroyemi, and E. McKinnon. 2012. 'Wellbeing and Civil Society: Estimating the Value of Volunteering Using Subjective Wellbeing Data'. Working Paper 112. London, UK: Department for Work and Pensions. https://www.gov.uk/government/publications/wellbeing-and-civil-society-estimating-the-value-of-volunteering-using-subjective-wellbeing-data-wp112.

Fujiwara, Daniel. 2013. 'Museums and Happiness: The Value of Participating in Museums and the Arts'. London, UK: The Happy Museum Project. http://media.wix.com/ugd/9ccf1d_35c943a360664351a930bc459891a060.pdf.

Fujiwara, Daniel, and Paul Dolan. 2015. 'Happiness-Based Policy Analysis'. In *Oxford Handbook of Wellbeing and Public Policy*, edited by M Adler and M Fleurbaey.

Fujiwara, Daniel, Ricky N. Lawton, and Susana Mourato. 2019. 'More than a Good Book: Contingent Valuation of Public Library Services in England'. *Journal of Cultural Economics* 43 (3):639–666. https://doi.org/10.1007/s10824-019-09369-w.

Goodspeed, T. 2017. 'An Evaluation of Social Return Using Willingness to Pay'. London, UK: morethanoutputs.

Gražulevičiūtė-Vileniškė, Indrė, Vytautas Janilionis, Jūratė Guščinskienė, and Ligita Azukaitė. 2011. 'Contingent Valuation of Built Heritage Properties in a Transition Country: A Case of Lithuania'. *International Journal of Strategic Property Management* 15 (4):393–415.

Greene, William H. 2007. 'Discrete Choice Modeling'. SSRN Scholarly Paper ID 985611. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=985611.

Groot, Rudolf de, Luke Brander, Sander van der Ploeg, Robert Costanza, Florence Bernard, Leon Braat, Mike Christie, et al. 2012. 'Global Estimates of the Value of Ecosystems and Their Services in Monetary Units'. *Ecosystem Services* 1 (1):50–61. https://doi.org/10.1016/j.ecoser.2012.07.005.

Haddaway, Neal Robert, Alexandra Mary Collins, Deborah Coughlin, and Stuart Kirk. 2015. 'The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching'. *PLOS ONE* 10 (9):e0138237. https://doi.org/10.1371/journal.pone.0138237.

Herrero, Luis César, José ángel Sanz, and María Devesa. 2011. 'Measuring the Economic Value and Social Viability of a Cultural Festival as a Tourism Prototype'. *Tourism Economics* 17 (3):639–53. https://doi.org/10.5367/te.2011.0057.

Hicks, J. R., and R. G. D. Allen. 1934. 'A Reconsideration of the Theory of Value. Part I'. *Economica* 1 (1):52–76.

Hicks, Robert L., and Bonnie M. Queen. 2016. 'Valuing Historical and Open Space Amenities with Hedonic Property Valuation Models'. *Agricultural and Resource Economics Review* 45 (1):44–67.

HM Treasury. 2011. 'The Green Book: Appraisal and Evaluation in Central Government'. HM Treasury. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf.

Horowitz, John K., and Kenneth E. McConnell. 2002. 'A Review of WTA/WTP Studies'. *Journal of Environmental Economics and Management* 44 (3):426–47. https://doi.org/10.1006/jeem.2001.1215.

Jaffry, Shabbar, and Alexandros Apostolakis. 2011. 'Evaluating Individual Preferences for the British Museum'. *Journal of Cultural Economics* 35 (1):49–75. https://doi.org/10.1007/s10824-010-9133-z.

Johnston, R., J. Rolfe, R.S. Rosenberger, and R. Brouwer. 2015. *Benefit Transfer of Environmental and Resource Values - A Guide for Researchers and Practitioners*. The Economics of Non-Market Goods and Resources 14. London, UK: Springer. http://www.springer.com/gb/book/9789401799294.

Johnston, R.J., K.J. Boyle, W Adamowicz, J. Bennett, R Brouwer, T.A. Cameron, W.M. Hanemann, et al. 2017. 'Contemporary Guidance for Stated Preference Studies'. Association of Environmental and Resource Economists.

Kahneman, Daniel, Edward Diener, and Norbert Schwarz. 2003. *Well-Being: Foundations of Hedonic Psychology*. Vol. 13. New York, NY: Russell Sage Foundation.

Kim, Soo-Il, and Timothy C. Haab. 2009. 'Temporal Insensitivity of Willingness to Pay and Implied Discount Rates'. *Resource and Energy Economics* 31 (2):89–102. https://doi.org/10.1016/j.reseneeco.2008.09.006.

Kim, Younjun, Catherine L. Kling, and Jinhua Zhao. 2015. 'Understanding Behavioral Explanations of the WTP-WTA Divergence Through a Neoclassical Lens: Implications for Environmental Policy'. *Annual Review of Resource Economics* 7 (1):169–87. https://doi.org/10.1146/annurev-resource-100913-012501.

Lawton, R., D. Fujiwara, S. Mourato, H. Bakhshi, A. Lagarde, and J. Davies. 2018. 'The Economic Value of Heritage: A Benefit Transfer Study'. Arts and Humanities Research Council.

Lawton, Ricky N., Susana Mourato, Daniel Fujiwara, and Hasan Bakhshi. 2019. 'Comparing the Effect of Oath Commitments and Cheap Talk Entreaties in Contingent Valuation Surveys: A Randomised Field Experiment'. *Journal of Environmental Economics and Policy*, November, 1–17. https://doi.org/10.1080/21606544.2019.1689174.

Lazrak, Faroek, Peter Nijkamp, Piet Rietveld, and Jan Rouwendal. 2014. 'The Market Value of Cultural Heritage in Urban Areas: An Application of Spatial Hedonic Pricing'. *Journal of Geographical Systems* 16 (1):89–114. https://doi.org/10.1007/s10109-013-0188-1.

Lemyre, X, J Mader, and M Ambard. 2018. 'Quantifying and Valuing the Wellbeing Impacts of Arts, Culture and Sports in Canada'. Canadian Heritage. https://cudo.carleton.ca/system/files/dli_training/4330/wellbeing-impacts-arts-culture-and-sports-canada-final.pdf.

Lundhede, Thomas, Trine Bille, and Berit Hasler. 2013. 'Exploring Preferences and Non-Use Values for Hidden Archaeological Artefacts: A Case from Denmark'. *International Journal of Cultural Policy* 19 (4):501–30.

Maddison, David, and Susana Mourato. 2001. 'Valuing Different Road Options for Stonehenge'. *Conservation and Management of Archaeological Sites* 4 (4):203–12. https://doi.org/10.1179/135050301793138182.

Marsh, K., S. MacKay, D. Morton, W. Parry, E. Bertranou, J. Lewsie, R. Sarmah, and P. Dolan. 2010. *CASE: Understanding the Drivers, Impact and Value of Engagement in Culture and Sport [Online]. Department for Culture, Media and Sport*.

Mazzanti, Massimiliano. 2003. 'Valuing Cultural Heritage in a Multi-Attribute Framework Microeconomic Perspectives and Policy Implications'. *The Journal of Socio-Economics* 32 (5):549–69. https://doi.org/10.1016/j.socec.2003.08.009.

Melstrom, Richard T. 2014. 'Valuing Historic Battlefields: An Application of the Travel Cost Method to Three American Civil War Battlefields'. *Journal of Cultural Economics* 38 (3):223–36. https://doi.org/10.1007/s10824-013-9209-7.

Mourato, S., E. Fimereli, D. Contu, C. Gaskell, and C. Boniatti-Pavese. 2014. 'The Economic Benefits of Cultural Built Heritage Interiors Conservation from Climate Change Damages in Europe'. WP6 Final Report. London, UK: Grantham Research Institute on Climate Change and the Environment. https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjAx_WBlOTNAhXmAMAKHX9cAJ8QFggeMAA&url=http%3A%2F%2Fwww.climateforculture.eu%2Findex.php%3Finhalt%3Ddownload%26file%3Dpages%2Fuser%2Fdownloads%2Fproject_results%2FD_06.1_final_publish.pdf&usg=AFQjCNEmpiNVVbd1Kr66uDCAHhWep4PYWw&sig2=vfs6IqUMCX-ccnmg1HWP7w.

Mwebaze, Paul, and Jeff Bennett. 2012. 'Valuing Australian Botanic Collections: A Combined Travel-Cost and Contingent Valuation Study*'. *Australian Journal of Agricultural and Resource Economics* 56 (4):498–520. https://doi.org/10.1111/j.1467-8489.2012.00595.x.

Noonan, Douglas S. 2003. 'Contingent Valuation and Cultural Resources: A Meta-Analytic Review of the Literature'. *Journal of Cultural Economics* 27 (3–4):159–76. https://doi.org/10.1023/A:1026371110799.

Ofcom. 2008. 'Assessing the Value of Public Service Programming on ITV1, Channel 4 and Five'. London, UK: Holden Permain. http://stakeholders.ofcom.org.uk/binaries/consultations/psb2_phase2/annexes/annex6.pdf.

Parsons, George R. 2017. 'Travel Cost Models'. In *A Primer on Nonmarket Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown, 187–233. The Economics of Non-Market Goods and Resources. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7104-8_6.

Pearce, David, and Ece O'zdemiroglu. 2002. 'Economic Valuation with Stated Preference Techniques'. London, UK: Department for Transport, Local Government and the Regions. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/191522/Economic_valuation_with_stated_preference_techniques.pdf.

Peng, Cong, Stephen Gibbons, and Cheng Keat Tang. 2019. 'Valuing the Environmental Benefits of Canals Using House Prices'.

Provins, Allan, David Pearce, Ece Ozdemiroglu, Susana Mourato, and Sian Morse-Jones. 2008. 'Valuation of the Historic Environment: The Scope for Using Economic Valuation Evidence in the Appraisal of Heritage-Related Projects'. *Progress in Planning* 69 (4):131–175.

Pung, Caroline, Ann Clarke, and Laurie Patten. 2004. 'Measuring the Economic Impact of the British Library'. *New Review of Academic Librarianship* 10 (1):79–102. https://doi.org/10.1080/13614530412331296826.

Reiter, Gisela, Nicole Gonser, Markus Grammel, and Johann Gründl. 2017. 'Young Audiences and Their Valuation of Public Service Media'. *Public Service Media in the Networked Society. Nordicom, Gotemburg*, 211–26.

Rouwendal, Jan, and Jaap Boter. 2009. 'Assessing the Value of Museums with a Combined Discrete Choice/Count Data Model'. *Applied Economics* 41 (11):1417–36.

Sanz, José Ángel, Luis César Herrero, and Ana María Bedate. 2003. 'Contingent Valuation and Semiparametric Methods: A Case Study of the National Museum of Sculpture in Valladolid, Spain'. *Journal of Cultural Economics* 27 (3–4):241–57. https://doi.org/10.1023/A:1026353218280.

Saz-Salazar, Salvador del, Ana Navarrete-Tudela, José Ramón Alcalá-Mellado, and Daniel Carlos del Saz-Salazar. 2017. 'On the Use of Life Satisfaction Data for Valuing Cultural Goods: A First Attempt and a Comparison with the Contingent Valuation Method'. *Journal of Happiness Studies*, November, 1–22. https://doi.org/10.1007/s10902-017-9942-2.

Snowball, J. D. 2005. 'Art for the Masses? Justification for the Public Support of the Arts in Developing Countries – Two Arts Festivals in South Africa'. *Journal of Cultural Economics* 29 (2):107–25. https://doi.org/10.1007/s10824-005-5064-5.

Stevenson, David. 2013. 'Reaching a "Legitimate" Value? A Contingent Valuation Study of the National Galleries of Scotland'. *Museum Management and Curatorship* 28 (4):377–93. https://doi.org/10.1080/09647775.2013.831250.

Throsby, David. 1994. 'The Production and Consumption of the Arts: A View of Cultural Economics'. *Journal of Economic Literature* 32 (1):1–29.

———. 2003. 'Determining the Value of Cultural Goods: How Much (or How Little) Does Contingent Valuation Tell Us?' *Journal of Cultural Economics* 27 (3–4):275–85. https://doi.org/10.1023/A:1026353905772.

Tourkolias, Christos, Theodora Skiada, Sebastian Mirasgedis, and Danae Diakoulaki. 2015. 'Application of the Travel Cost Method for the Valuation of the Poseidon Temple in Sounio, Greece'. *Journal of Cultural Heritage* 16 (4):567–74. https://doi.org/10.1016/j.culher.2014.09.011.

Willis, Kenneth G. 2009. 'Assessing Visitor Preferences in the Management of Archaeological and Heritage Attractions: A Case Study of Hadrian's Roman Wall'. *International Journal of Tourism Research* 11 (5):487–505.

Willis, K.G. 2014. 'The Use of Stated Preference Methods to Value Cultural Heritage'. In *Handbook of the Economics of Art and Culture, Volume 2*, by V.A. Ginsburgh and D. Throsby, 145–81. London, UK: Elsevier. http://dx.doi.org/10.1016/B978-0-444-53776-8.00007-6.