Bayesian methods in palliative care research

Parker, Richard A.; Sande, Tonje A.; Laird, Barry; Hoskin, Peter; Fallon, Marie; Colvin, Lesley

Published in:
BMJ Supportive and Palliative Care

DOI:
10.1136/bmjspcare-2019-002160

Publication date:
2020

Document Version
Peer reviewed version

Experience of using Bayesian methods in Palliative Care Research: an example in cancer induced bone pain (CIBP)

Richard A. Parker[1]*, Tonje A. Sande[1]*, Barry J. A. Laird[2], Peter Hoskin[3], Marie T. Fallon[2]**, Lesley Colvin[4]**

*Joint first authors
**Joint senior authors

Affiliations:

[1] Usher Institute, University of Edinburgh, Scotland, United Kingdom
[2] Institute of Genetics and Molecular Medicine, University of Edinburgh, Scotland, United Kingdom
[3] Mount Vernon Cancer Centre and University of Manchester, United Kingdom
[4] Division of Population Health and Genomics, University of Dundee, Scotland, United Kingdom

Corresponding author:

Richard A. Parker,
Edinburgh Clinical Trials Unit,
Usher Institute,
The University of Edinburgh
Level 2, NINE,
9 Little France Road,
Edinburgh BioQuarter,
Edinburgh
EH16 4UX
Richard.parker@ed.ac.uk
0131 651 9953

Word Count: 2,476

This article has been accepted for publication in BMJ Journal of Supportive and Palliative Care (2020) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/bmjspcare-2019-002160.

ABSTRACT

Objective

To show how a simple Bayesian analysis method can be used to improve the evidence base in patient populations where recruitment and retention are challenging.

Methods

A Bayesian conjugate analysis method was applied to binary data from the Thermal testing in Bone Pain (TiBoP) study: a prospective diagnostic accuracy/predictive study in patients with cancer-induced bone pain (CIBP). This study aimed to evaluate the clinical utility of a simple bedside tool to identify who was most likely to benefit from palliative radiotherapy (XRT) for CIBP.

Results

Recruitment and retention of patients was challenging due to the frail population, with only 27 patients available for the primary analysis. The Bayesian method allowed us to make use of prior work done in this area and combine it with the TiBoP data to maximise the informativeness of the results. Positive and negative predictive values were estimated with greater precision, and interpretation of results was facilitated by use of direct probability statements. In particular, there was only 7% probability that the true positive predictive value was above 80%.

Conclusions

Several advantages of using Bayesian analysis are illustrated in this article. The Bayesian method allowed us to gain greater confidence in our interpretation of the results despite the small sample size by allowing us to incorporate data from a previous similar study. We suggest that this method is likely to be useful for the analysis of small diagnostic or predictive studies when prior information is available.

Keywords: Diagnostic accuracy, conjugate, sensitivity, specificity, Bayesian analysis, beta distribution, small sample size

INTRODUCTION

The TiBoP study was concerned with improving outcomes for patients with cancer-induced bone pain (CIBP). CIBP is a consequence of metastases to bone; and can have a major impact on day-to-day function and quality of life.[1] Currently, the gold standard treatment is palliative radiotherapy (XRT), although only approximately half of patients will achieve satisfactory pain relief, and this may take up to six weeks to work properly.[2,3]

Somatosensory testing, used to define pain mechanisms in individual patients, has shown some promise in predicting treatment response in neuropathic pain.[4] Our previous pilot work demonstrated sensory changes in CIBP, with alterations in skin sensation overlying the area of CIBP.[1] This pilot work suggested that altered thermal sensitivity on the skin overlying the site of painful bone metastases might have value in predicting an increased likelihood of a good outcome from XRT.[5]

Therefore, the Thermal testing in Bone Pain (TiBoP) study was carried out to assess the performance of a simple thermal sensitivity measure that could be used by non-specialists in the community, to identify who was most likely to get analgesic benefit from XRT for

CIBP. The study faced challenges with recruitment and retention of patients, and the final study sample size was small. Conducting research in palliative care can be challenging due to the frailty of the patient population making it difficult to establish a robust evidence base. There is need for using innovative methods to deal with this challenging research environment.

This short report shows how a simple Bayesian analysis can be used to maximise the value of small diagnostic studies by allowing previous data to bolster the results.

METHODS

The TiBoP study

The TiBoP study was a prospective exploratory study, carried out in two centres (Edinburgh and Dundee) and approved by South Central - Oxford C Research Ethics Committee No:16/SC/0260. All patients gave written informed consent to take part in the study.

The thermal sensitivity tool evaluated in this study involved using warm (40°C) and cool (25°C) thermal rollers (Rolltemp, Somedic, Sweden, CE marked) to assess the thermal sensitivity of the skin overlying the painful bone metastasis in comparison to a corresponding unaffected (control) area.

Eligible patients were adults (aged 18 or older) scheduled for palliative XRT for treatment of CIBP. A convenience sample of eligible patients were recruited between October 2016 and May 2018; and all were tested using the thermal sensitivity tool prior to receiving XRT. The primary endpoint was pre-specified to be worst pain score at six weeks post-XRT, using the Brief Pain Inventory (BPI) questionnaire. Specifically, the primary endpoint was defined as either (i) a 30% or higher reduction in worst pain score (Q3 of the BPI questionnaire), or (ii) a worst pain score of zero at six weeks. This mirrored the endpoint for pain response used in our previous study.[5,6] Our hypothesis was that patients experiencing "abnormal sensitivity" based on the thermal sensitivity test were more likely to achieve a response to XRT (i.e. pain reduction).

Statistical methods

The statistical analysis was concerned with making inference about the true values of the sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) of the thermal sensitivity score with respect to experiencing a response to XRT. In practice, there were two components to the thermal sensitivity test: a test involving a warm roller and a separate test involving a cool roller. If patients reported abnormal sensitivity for both tests then the overall score was assumed to be "abnormal". A separate analysis was conducted under the assumption that *any* of the two tests needed to be "abnormal" for the overall score to be "abnormal".

The primary analysis was conducted in a Bayesian framework, using Bayes' theorem.[7-9] The key to understanding Bayesian analysis is that we begin with prior information

regarding the parameters of interest (e.g. sensitivity and specificity); and then we use Bayes' theorem to update our prior information based on observed data.[9] In the case of the TiBoP study, prior information about the likely value of the proportions of interest was gathered from a previous pilot study conducted as part of an MD thesis.[6] In Bayesian analysis, prior information or beliefs are usually expressed as a range of possible values through specification of a probability distribution.[9] In our case, a beta distribution was used for the prior and binomial distribution to model the binary thermal sensitivity test data (e.g. "abnormal" response or not).

After combining the model for the observed data with the model for the prior information using Bayes' theorem, we obtain a "posterior distribution", which gives us a probability distribution for the probability of the proportions given the observed data, which is what we are really interested in. In our case, we get a posterior distribution that has a beta distributional form just like our prior distribution. Formally, this is called a "conjugate analysis" [7] and we say that the beta distribution is "*conjugate"* to the binomial distribution.

R software [10] was used to perform the analysis. Graphs of posterior distributions were generated for all diagnostic test statistics of interest (i.e. probability distributions for the true parameters of interest: NPV, PPV, sensitivity, specificity etc.), while posterior means and 95% highest posterior density (HPD) credible intervals were calculated to show the likely range of values for the true parameter (e.g. true NPV).

A specific (informative) prior was pre-specified based on the previous pilot study, but we also checked the sensitivity of the results to this prior by using a (i) a weakly informative prior and (ii) a flat completely uninformative prior (Beta(1,1)). This allowed us to compare our results with models for which the observed TiBoP study data dominated.

Results were compared to the classical (frequentist) approach of calculating 95% confidence intervals around parameters without utilizing prior data.

Further details of the statistical analysis are provided in the online supplementary file along with a mini-literature review suggesting that the use of this method is very uncommon in practice.

RESULTS

Forty patients were recruited to the study between October 2016 and April 2018 from two locations (34 from Edinburgh and six from Dundee, United Kingdom). Twenty-seven patients (67%) completed the primary outcome assessment at six weeks.

Of the 27 patients recording primary outcome data, the mean age was 65 (SD 9.5, range 43 to 84). Eleven patients (41%) were female. Thirteen had a primary diagnosis of prostate cancer (48%), eight had breast cancer (30%) and the remaining six patients (22%) had various other types of cancer.

Considering the comparison of patients with both abnormal thermal sensitivity tests compared to those with at least one normal, the observed sensitivity, specificity, PPV and

NPV of the thermal sensitivity score (with corresponding exact binomial 95% confidence intervals) were calculated as 9/15 (60%, 95% CI 32% to 84%), 5/12 (42%, 95% CI 15% to 72%), 9/16 (56%, 95% CI 30% to 80%), and 5/11 (45%, 95% CI 17% to 77%) respectively. These 95% confidence intervals were computed using the standard classical method ignoring prior data.

The observed results with classical confidence intervals suggest that thermal sensitivity score is a poor predictor of positive response to XRT. PPV and NPV are close to 50% and specificity is very low. Confidence intervals were very wide, so there was a great deal of uncertainty associated with the estimates when just considering the current study data.

After using results from the previous pilot study to inform the prior distribution, Bayes' Theorem was used to produce plots of the posterior distributions for each diagnostic test statistic (see Figure 1).

The posterior mean PPV (95% credible interval) was 70% (57% to 83%), suggesting that it is unlikely that the true PPV for the thermal sensitivity tool is above 83%. Indeed, the probability that the true PPV is above 80% was only 7% ($\mathbb{P}(PPV > 0.80) = 0.07$). This means that the thermal sensitivity test is unlikely to be useful in accurately identifying patients who will go on to get a positive response to XRT at six weeks. The credible interval upper bound is similar to the classical frequentist confidence interval of 80%, but note that the interval is much narrower since we have combined with the previous data (PPV estimated as 81%) to increase the precision of estimation. Using a flat non-informative prior

(i.e. ignoring the prior data we have), results in a credible interval from 33% to 77%, which is similar to that from the frequentist 95% CI as we might expect.

For NPV, the posterior mean was 48%, but the 95% credible interval had a very wide range from 30% to 66% due to the low number of patients in this category. Note that this interval is also much narrower than the corresponding frequentist 95% confidence interval of 17% to 77%. Indeed, it is true in general that precision of estimation will often be improved through using Bayesian methods, particularly if specific informative priors are used and the study sample size is small.

To provide a more extreme example: only 3 patients had thermal test results which were both normal. Two of these did not show a positive response to XRT, and so the NPV under the "at least one abnormal" classification was calculated as 2/3 (67%). Naturally, the standard 95% CI for the NPV was extremely wide (9% to 99%). However, after combining with the prior information (NPV 3/4, 75%), the 95% HPD interval was 38% to 94%, which although still wide, does inform us that very high values of the NPV above 94% are unlikely. We can also calculate $\mathbb{P}(NPV > 0.90) = 0.04$ which supports this conclusion.

In contrast, the PPV ("at least one abnormal" classification), was based on more substantial sample sizes (PPV was 14/24 (58%) for the current study and 28/38 (74%) in the previous study). In this example, the 95% HDP interval was 56% to 78% compared to a 95% CI of 57% to 87%. Thus, our interval upper bound reduces from 87% down to 78% with the addition of prior information. We can also calculate $\mathbb{P}(PPV > 0.80) = 0.01$, which shows there is only 1% probability that the true PPV is above 80%.

For the above examples, the estimates based on the prior information are not too inconsistent with those from the present study. If hypothetically, the prior PPV was only 3% (1/38), then the 95% HPD interval becomes substantially different, 15% to 36%, albeit the interval is still much narrower than the corresponding 95% CI.

The Bayesian results suggested that the thermal sensitivity tool alone is unlikely to be useful in practice for identifying patients who experience a response to XRT treatment. This was despite the use of an informative prior distribution based on promising results from an earlier study.[5,6]

Full analysis results are provided in a supplementary file.

DISCUSSION

Bayesian analysis has some advantages over classical analysis. In particular:

(i) It makes full use of previous work done in the same area so that it informs the statistical analysis. The Bayesian argument is that no study happens in isolation, and that it makes sense to incorporate external information when performing statistical inference because scientific progress generally always involves building on what has been done before.

(ii) In small studies, the informativeness of the results can be maximised through more precise estimates of diagnostic test measurements (e.g. NPV and PPV).

(iii) Interpretation of results is easier and more intuitive. For example, the true value lies within a 95% credible interval with 95% probability.

(iv) Posterior distribution graphs of parameters can be generated from Bayesian analysis (see Figure 1) to provide helpful visual information regarding the likely true value.

(v) In addition, direct probability statements can be made that are easier to interpret clinically.[11] As shown above, we can easily answer questions such as "What is the probability that the true negative predictive value is above 80%?" Whereas in classical analysis it is very difficult to answer such questions.

(vi) Bayesian analysis is more ethical because it fully exploits the clinical experience of past patients to maximise the potential of a small sample size to generate meaningful results.[11] Even data from very small prior studies is not wasted and can contribute to the analysis.

Bayesian analysis was particularly useful for the TiBoP example because data was available from a very similar previous study, the study was small (and so we could take full advantage of the improvement in precision resulting from incorporating prior data), and there was an ethical imperative to make maximal use of all data collected. These advantages more generally apply in studies in palliative care, where patients are frail, with life limiting disease and therefore it is especially important to ensure that precious data collected from patients is not wasted. In general patients are supportive of participating in clinical research, often for altruistic reasons, although naturally there may be burden placed on these patients when collecting data.[12]

In the TiBoP study, there was a consistent gain in precision from using Bayesian methods: our 95% credible intervals were narrower than the corresponding 95% confidence intervals.

In some situations (e.g. in our NPV example), this enabled us to salvage data that may otherwise have been completely unusable due to the tiny sample size. However, there is a note of caution associated with this. As we saw in the PPV example, artificially changing the prior information led to a dramatic change in the values of the credible interval estimates. This was because we were using a specific informative prior to combine with the observed information, which places a high weight on prior information. This was justified in our case since the studies were very similar in design, with the same lead researchers and assessment approaches used for both studies, and the majority of patients recruited from the same centre (Edinburgh). However, it was still important for us to test the sensitivity of the results to the use of non-informative priors (see supplementary file).

If on the other hand, the previous study was conducted under very different conditions, had a different patient population, or was more susceptible to bias, then less weight should have been placed on the prior information and it would have been necessary to use vague or non-informative priors for our primary analysis. However, collaterally we lose the advantage of improved precision from using Bayesian methods.

Bayesian analysis may be less useful in circumstances which nullify some of the advantages listed above. For example, if our study has a large sample size with no similar previous studies, then finding suitable information to inform the prior distribution may be difficult and there may be little or no gain in precision from using a Bayesian approach. Nevertheless, some advantages of Bayesian analysis will still remain regardless of the context (e.g. the ability to make direct probability statements).

The methodology used in this study is particularly beneficial in settings where it is difficult to establish a robust evidence base (e.g. in frail populations or rare conditions) due to its ability to effectively assimilate prior data and enhance the value of information from small studies.

Acknowledgements

The previous pilot study data used in this study was collected by Angela Scott for her MD thesis: "Cancer-Induced Bone Pain (CIBP): Clinical Characterisation and Biomarker Development" (http://hdl.handle.net/1842/24294), funded by the Translational Medicine Research Collaboration (TMRC). RP was partly supported in this work by NHS Lothian via the Edinburgh Clinical Trials Unit. PJH is supported by the NIHR Manchester Biomedical Research Centre.

Competing Interest

Competing Interest: None declared.

Funding

The TiBoP study was funded by Marie Curie [MC grant reference MCCC-RP-15-A19005PI].

Data

Requests for data sharing should be directed to: ECTUdatashare@ed.ac.uk

REFERENCES

[1] Laird BJ, Walley J, Murray GD, et al. Characterization of cancer-induced bone pain: an exploratory study. *Supportive Care in Cancer* 2011; 19: 1393-401.

[2] Pin Y, Paix A, Le Fèvre C, et al. A systematic review of palliative bone radiotherapy based on pain relief and retreatment rates. *Crit Rev Oncol Hematol* 2018; 123: 132-13.

[3] Chow R, Hoskin P, Chan S, et al. Efficacy of multiple fraction conventional radiation therapy for painful uncomplicated bone metastases: A systematic review. *Radiother Oncol* 2017; 122(3): 323-33.

[4] Demant DT, Lund K, Vollert J, et al. The effect of oxcarbazepine in peripheral neuropathic pain depends on pain phenotype: a randomised, double-blind, placebo-controlled phenotype-stratified study. *PAIN®* 2014; 155(11): 2263-73.

[5] Scott AC, McConnell S, Laird B, et al. Quantitative Sensory Testing to assess the sensory characteristics of cancer-induced bone pain after radiotherapy and potential clinical biomarkers of response. European journal of pain 2012; 16(1): 123-33.

[6] Scott AC. *Cancer-Induced Bone Pain (CIBP): Clinical Characterisation and Biomarker Development.* MD Thesis, University of Edinburgh, UK, 2010. (http://hdl.handle.net/1842/24294, 2010, accessed 9 September 2019)

[7] Spiegelhalter DJ, Abrams KR, and Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons, 2004.

[8] Abrams KR. Bayes' theorem. In: Everitt BS and Palmer CR (eds) *Encyclopaedic Companion to Medical Statistics.* Hodder Arnold, 2005, pp. 22-23.

[9] Kirkwood BR and Sterne JA. *Essential medical statistics.* Second edition. Blackwell Science Ltd., 2003, pp. 388-389.

[10] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL { HYPERLINK "https://www.R-project.org/" }.

[11] Abrams KR. Bayesian Methods. In: Everitt BS and Palmer CR (eds) Encyclopaedic Companion to Medical Statistics. Hodder Arnold, 2005, pp. 23-27.

[12] Todd AM, Laird BJ, Boyle D, et al. A systematic review examining the literature on attitudes of patients with advanced cancer toward research. *J Pain Symptom Management* 2009; 37(6): 1078-85.

FIGURE

Figure 1: Plots showing the posterior distributions for the diagnostic test parameters under the strategy of using "at least one test abnormal" as the diagnostic test marker to predict positive response. Solid line indicates specific prior, dashed line is weakly specific prior, and dotted line is uninformative prior.