

RESEARCH

Open Access



Survival estimates stratified by the Nottingham Prognostic Index for early breast cancer: a systematic review and meta-analysis of observational studies

Ewan Gray^{1*}, Anna Donten², Katherine Payne² and Peter S. Hall¹

Abstract

Background: Estimates of survival for women diagnosed with early staged breast cancer are available based on stratification into prognostic categories defined using the Nottingham Prognostic Index (NPI). This review aimed to identify and summarize the estimated survival statistics from separate sources in the literature and to explore the extent of between-study heterogeneity in survival estimates.

Methods: Observational studies in women diagnosed with early and locally advanced breast cancer reporting overall survival by NPI category were identified using a systematic literature search. An exploratory meta-analysis was conducted to describe survival estimates and assess between-study heterogeneity.

Results: Twenty-eight studies were identified. Nineteen studies with sufficient data on overall survival were included in meta-analysis. A high level of heterogeneity in survival estimates was evident with I^2 values in the range of 90 to 98%.

Conclusions: The substantial differences between studies in the relationship between NPI categories and survival at 5 and 10 years poses challenges for use of this prognostic score in both clinical settings and in decision-analytic model-based economic evaluations.

Keywords: Breast cancer, Prognosis, Precision medicine, Meta-analysis, Nottingham Prognostic Index

Background

The Nottingham Prognostic Index (NPI) is a commonly used, clinically relevant and internationally validated [1, 2] system for classifying early and locally advanced breast cancer cases (TNM stages I, II, and III [3]) into three or more prognostic groups [4, 5]. Although more advanced models [6, 7] have superseded NPI in some applications, it continues to have an active role in clinical practice and research [8]. In clinical practice, NPI and related prognostic models have an important application for patients and clinicians to inform the decision of whether or not to undergo adjuvant chemotherapy following surgery [9], an early example of what is commonly called precision medicine [10].

The NPI can be used to provide prognostic information by assigning individuals into prognostic categories and then applying the survival estimates from a previous cohort study [11]. As well as providing prognostic estimates for individual patients, the NPI is useful in the context of economic evaluations. In economic evaluation of new interventions, overall survival is a key input for decision-analytic models designed to quantify the incremental costs and benefits and inform if, and how, to allocate finite healthcare budgets towards new screening and management options [12]. Decision-analytic models have a key role in the evaluation of new technologies when clinical trials are not likely to be feasible or timely in terms of producing robust evidence of the impact of introducing new programmes or changing existing programmes [13].

The NPI categories are a linear combination of three prognostic factors: tumour size (maximum diameter in

* Correspondence: ewan.gray@ed.ac.uk

¹The University of Edinburgh, Edinburgh, UK

Full list of author information is available at the end of the article



millimetres), histological grade (1- to 3-point scale) and lymph node staging (1- to 3-point scale). Standardized published cut-offs (see Additional file 1: Appendix 1) are used to form prognostic categories within a population (e.g. good, moderate and poor prognostic groups).

This study aimed to identify and describe all published observational studies reporting NPI-category-specific overall survival following a diagnosis of early and/or locally advanced breast cancer. A meta-analysis was used to explore the between-study heterogeneity in survival estimates.

Methods

A systematic review and meta-analysis, following published recommendations [14, 15], were used to identify all published cohort studies investigating the survival of early and locally advanced breast cancer patients stratified into prognostic groups using the NPI and to synthesize the results. This systematic review was not registered with PROSPERO.

Data sources and searches

Two databases (Embase: date of inception 1974 to 9 November 2016; MEDLINE: date of inception 1946 to 9 November 2016) were searched using bespoke electronic search strategies (see Additional file 1: Appendix 2) informed by search criteria designed by Nelson et al. [16]. Hand searching of reference lists of included studies was also undertaken.

Study selection

Retrieved titles and abstracts were screened independently by two reviewers (EG, AD) in accordance with the inclusion and exclusion criteria. The review includes all cohort studies¹ of women with early or locally advanced breast cancer, reporting all-cause mortality (overall survival) stratified by NPI group. Only journal articles or reports published in English were included. Studies were excluded if they were limited to only a specified sub-group of the full population (e.g. HER-2-positive patients only) or if they included patients with recurrent cancer² or *ductal carcinoma in situ*.³ To be included in the meta-analysis, a study must have reported survival estimates as tables of survivors/events per year or graphically as survival curves.

Quality assessment and data extraction

The Critical Appraisal Skills Programme (CASP) Cohort Study Checklist [17] was used to appraise the reporting quality of the selected studies (see Additional file 1: Appendix 3). The following data were extracted and tabulated by two reviewers: study setting, sample description, methods of analysis and survival estimates ('overall survival'). Overall survival data were presented

in identified studies as tables or graphs. Survival data from the studies using graphs were extracted using visual assessment. Guyot et al. suggest the use of algorithms for extracting data from digitized curves and conversion to time-to-event data using inverted Kaplan-Meier equations. Guyot et al.'s method was not used in this study because it requires more detailed reporting of numbers at risk (at multiple time points) than what was generally available in the included studies. Annual survival estimates and initial sample sizes were used recorded (see Additional file 1: Appendix 4 for further details of data extraction). Reconstruction of individual level time-to-event data from published survival curves has been suggested to improve the precision of meta-analysis of survival estimates. The advantages are that is potentially possible to include information regarding the censoring of observations [18] and combine survival data reported over differing lengths of follow-up.

Data analysis

All statistical analyses were conducted using Stata version 14. Forest plots were used to present individual study survival estimates for both 5 years and 10 years of follow-up. These are ordered from most distant to most recent year of mid-point of data collection. Meta-analysis, using both fixed and random effects [19] models, produced pooled estimates of both 5-year and 10-year survival, for each NPI category separately, and allowed calculation of between-study heterogeneity statistics.

Heterogeneity between-study included datasets was assessed primarily using the I^2 statistic because the interpretation is appropriate and useful from an exploratory and hypothesis generating perspective [20]. This statistic is an approximate measure of what part of the variance between the estimated effects in the meta-analysis is caused by study heterogeneity rather than sampling error. An I^2 score over 75% is typically taken to indicate high heterogeneity. The statistic describes the observed heterogeneity and should not be used to make inference regarding the range of true effects [21]. The Q statistic provides a statistical test of the null hypothesis that studies are homogenous (all estimating the same true effect). However, it should be noted that statistical test may be an underpowered test for heterogeneity when the number of studies is small [22].

The tau statistic was also considered, which provides an estimate of the between-study standard deviation in the true effects within a random effects meta-analysis framework [23]. The choice between using the fixed effects or the random effects meta-analysis depends on the assumed perspective regarding the underlying true effect(s) and the desired interpretation of the precision weighted average survival estimate [24]. Assumptions

about the underlying true effects for these data are discussed in the terminology of Spiegelhalter et al. [25]: ‘identical parameters’, ‘independent parameters’ or ‘exchangeable parameters’. The ‘identical parameters’ assumption is probably unrealistic in this setting given the diversity in study circumstances. Assuming ‘independent parameters’ may be reasonable if the variation between studies is caused by a variety of unique differences in specific settings, and the ‘exchangeable parameters’ assumption may also be reasonable if the variation between studies can be described by a single mixing distribution. An independent parameters assumption would motivate a fixed effects meta-analysis, while exchangeable parameters would motivate a random effects meta-analysis. In this exploratory study, we are ambivalent about which of these assumptions is the most appropriate, believing both may be reasonable, and therefore present both fixed and random effects estimates.

Results

A total of 28 studies were suitable for inclusion in the review (see Fig. 1) and 19 of these studies were included in the meta-analysis (see Table 1). Of the 28 studies, two studies [2, 26] provided the results of survival analysis for two data series; therefore, a total of 30 datasets were

included in the review and 20 in the meta-analysis. Three studies [27–29] provided only data up to 5-year survival and were excluded from the meta-analysis of 10-year survival. Nine studies (containing ten datasets) were excluded from any kind of meta-analysis due to the following reasons. The quality of the graphical presentation in two studies [30, 31] did not allow for extraction of the survival data. Three further studies [2, 32, 33] did not report numbers of patients in each NPI category. One [34] used different cut-off points for NPI that meant the categories were not comparable with the majority of the studies. Two studies [35, 36] reported insufficient data. One study [37] excluded patients who died in the first months after the surgery, and therefore, the provided survival estimates were incomplete. Eight data series from seven studies [1, 5, 26, 38–41] reported results from the original Nottingham case series and/or West Midlands Cancer registry (to which Nottingham contributes). These studies report partially overlapping cohorts of patients (see Fig. 2). To minimize ‘double-counting’, four studies were excluded from the meta-analysis which meant that the remaining included studies [1, 5, 41] contained minimal overlap in the reported cohorts. Among the remaining studies, Blamey et al. [1] does overlap with Allgood et al. [41]; however, the proportion of data that was shared was a relatively small proportion of

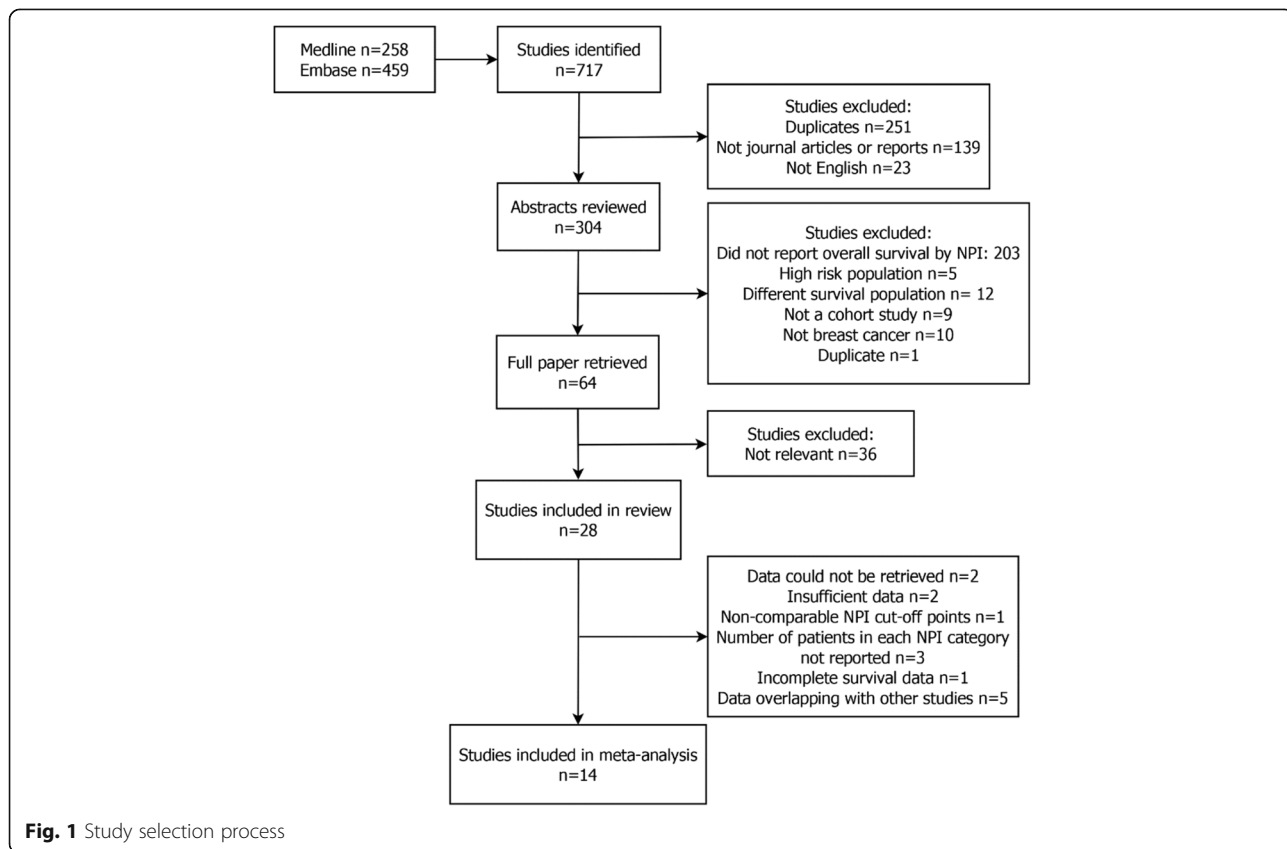


Fig. 1 Study selection process

Table 1 Summary of studies included in the review and meta-analysis

Study details		
Characteristic	Systematic review Number of data series ($n = 30$)	Meta-analysis Number of data series ($n = 14$)
Location		
UK only	14	6
Europe	10	5
USA	1	0
Other	5	3
Size of sample		
< 100 women	2	1
100 to 1000 women	10	6
1000 to 10,000 women	15	5
> 10,000 women	3	2
NPI categories		
Three	18	8
Five	2	2
Other	10	4
Length of analysis		
5 years	4	2
Between 5 and 10 years	3	1
10 years or more	23	11
Type of analysis		
Parametric	4	0
Cox proportional hazard model	20	9
Other (Kaplan-Meier estimator)	6	5

Note: One study [1] reported data from ten European countries including the UK; therefore, it is classified as 'European' study, not 'UK only'

the data because the first study reports a pooled European cohort of which approximately 15% of the data is from Nottingham, while the second study reports data from the West Midlands Cancer Registry to which Nottingham contributes only a minority of cases.

Description of studies

The earliest data reported in the published studies came from 1970 [42] and most recent data from 2014 [43] (see Fig. 2). Fourteen of the identified data series (47%) reported UK data. Ten studies reported data from other European countries, and one study [1] reported data from ten European countries, including the UK. One study [2] compared two datasets: from Finland and the USA. Single studies reported data from each of Canada [42], Japan [35], South Korea [32], Hong Kong [27] and Iran [43]. Study reporting

quality was generally assessed as satisfactory, and risk of bias in relation to the study survival estimates was considered to be low for all included studies (see Additional file 1: Appendix 3).

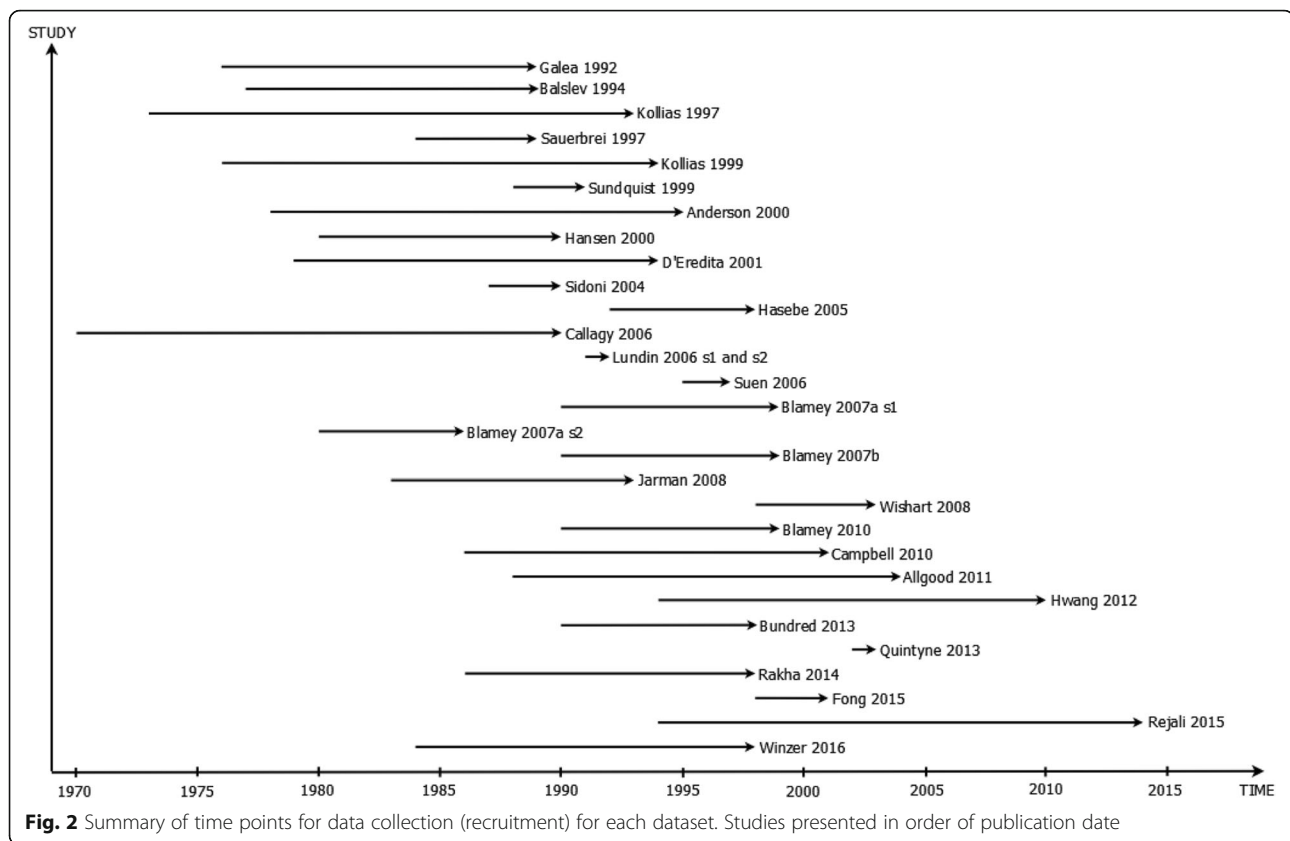
Start of each line represents first time point of data collection, and arrow head represents last time point of data collection for each study dataset.

Systematic review of evidence

There was diversity between the included studies in how women with breast cancer were classified by NPI category. The majority of the studies ($n = 18$; 60%) used three NPI-categories. Two studies [29, 44] (7%) grouped the sample into five NPI categories. The remaining ten (33%) studies used other numbers of NPI groups ranging from four to ten (see Table 1). Given the majority of studies used three NPI categories, the data from all studies were aggregated into three groups collapsing NPI categories where necessary. Three studies [30, 42, 45] did not provide cut-off points for NPI categories, and it was, therefore, assumed these studies used the standard cut-off values: NPI less than 3.4 for 'good prognostic group', between 3.4 and 5.4 for 'moderate prognostic group' and greater than 5.4 for 'poor prognostic group'. Two studies [28, 41] provided different cut-off points. It was assumed that observations with NPI less than 3.41 belong to the good prognostic group and observations with NPI less than 5.41 in the Jarman et al. [28] study and NPI less than 5.28 in Allgood et al. [41] study belong to the moderate prognostic group.

Figures 3 and 4 report the percentage of women surviving with breast cancer at two time points, 5 years and 10 years respectively, for the available datasets from which at least one of these statistics were available (Additional file 1: Appendix 5 for numbers in each study). One study [27] only reported survival data up to 9 years. Three studies [29, 40, 41] only provided survival data for the final year of analyses in each respective study.

The forest plots provide a clear visualisation of the trends between NPI categories within studies and, as these are ordered by the mid-point of data collection, differences in overall survival across time periods. The observed pattern of overall survival by NPI category was consistent across all datasets. Overall survival was seen to be noticeably worse in NPI category 3 compared with NPI categories 1 and 2. Visual inspection suggests substantial heterogeneity was observed in overall survival between the available datasets. To illustrate, 5-year survival for NPI category 1 was as low as 81.25% [27] but also as high as 99.26% [8]. Similarly, for NPI category 3, the lowest 5-year overall survival was 24.56% [5] and the highest was 83.55% [46].



Meta-analysis

The results of the meta-analysis are presented in Table 2 for both 5-year survival and 10-year survival statistics. There was evidence of considerable heterogeneity in the estimated survival statistics between the included datasets. The I^2 statistic indicated high heterogeneity for each of the NPI category with approximately 90 to 98% of the between-study variation estimated as being due to heterogeneity rather than sampling variation, suggesting pooling of survival estimates was inappropriate in all NPI categories. Heterogeneity was also considered present based on Q statistics which were statistically significant in all instances. Tau statistics provide further evidence for substantial heterogeneity with the reported standard deviations of effect size being 2, 7 and 18 percentage points (categories 1 to 3 respectively) for 5-year survival and 5, 11 and 14 percentage points for 10-year survival.

Discussion

This review identified a substantial number of observational studies that aimed to generate estimates of the overall survival of women diagnosed with early and/or locally advanced breast cancer stratified by prognostic category using NPI.

The strengths of this study include using a comprehensive, systematic review of the literature underpinned by robust review methods (see Additional file 2: PRISMA checklist). A broad electronic search strategy, supplemented with hand searching of reference lists of identified studies, to collate all the relevant studies and datasets. The search strategy and inclusion criteria achieved good sensitivity in identifying relevant studies. The identified studies included patient cohorts from a broad range of time periods for data collection, countries and age groups. Critical appraisal following a standard protocol found study reporting quality to be generally satisfactory. A number of relevant identified studies had to be excluded from the meta-analysis of overall survival as the relevant data could not be extracted from the published manuscript and authors were not able to provide the data required. The studies excluded at this stage did not appear to suggest different outcomes for survival compared to the included studies based on the more limited information available. We focussed on cohort studies and did not include randomized controlled trials (RCT). RCT studies were considered unlikely to report NPI-stratified survival estimates and are likely to have applied more extensive exclusion criteria than cohort studies. No formal quantitative assessment of publication bias was made. It is not clear in this context the

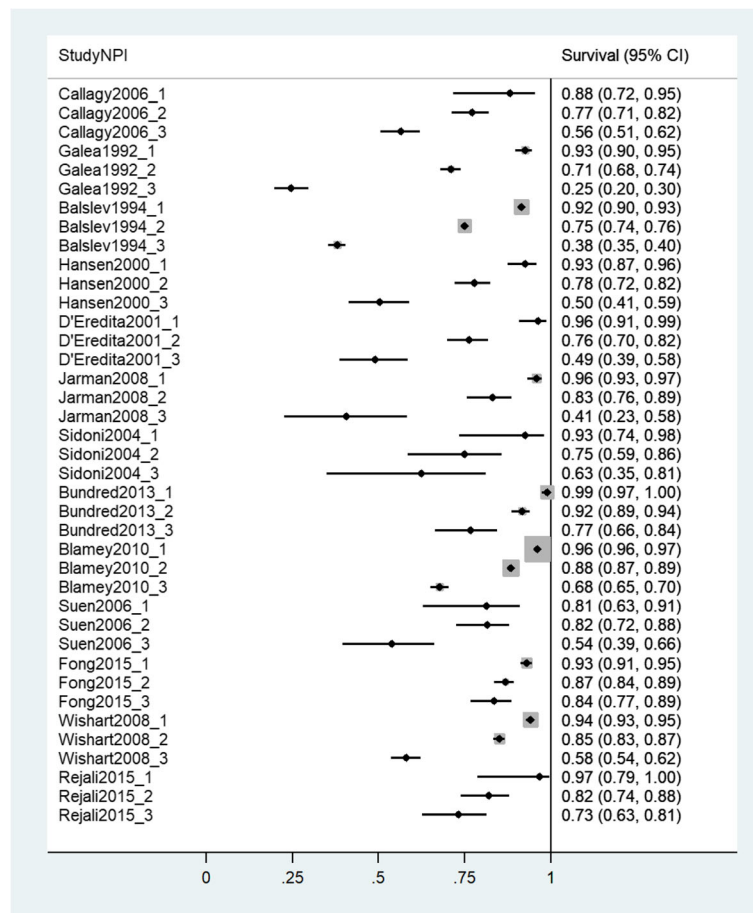


Fig. 3 Forest plot of 5-year survival estimates from individual studies. Studies presented in order of mid-point of data collection earliest to most recent. X-axis shows proportions of sample surviving at the specified time point

mechanism or result of publication bias that should be expected because each study was testing different hypothesis/hypotheses and in which the data extracted for this study were presented descriptively.

A particular concern in evidence synthesis is whether it was valid to pool data across multiple studies [20]. This exploratory meta-analysis investigated study heterogeneity and demonstrated that this was too severe for simple pooling of survival estimates to be appropriate for the purposes of estimating survival in the current population. It is of interest to understand the factors driving the heterogeneity in survival estimates; however, these data are insufficient to allow a robust meta-regression or multilevel analysis to test hypotheses about such determinants. There are only a small number of studies and many possible explanatory variables, including some for which only weak surrogate markers would be available from study level data. Investigation of the sources of between-study heterogeneity could be best achieved using an individual patient data meta-analysis to explore factors such as the

year of diagnosis of patients, country, screen detection rate and age.

This study had a number of limitations. To provide quantitative estimates of overall survival by NPI category, given the relative age of some of the identified studies, it was necessary to collate some data by visual inspection of published survival curves. This process will have introduced some measurement error. Due to lack of detailed reporting of losses to follow-up it was also necessary to assume that all studies had no loss to follow-up, which could result in different degrees of selection bias in the results if there were differential losses to follow-up across studies. Based on the published study protocols and the routine and often statutory collection of mortality data, we believe that losses to follow-up are likely to be minimal. Due to limitations of available resources, non-English language studies and those with insufficient reporting of data were excluded. Furthermore, it was considered infeasible based on available resources and likely data sharing restrictions, to attempt to collate individual patient data to address

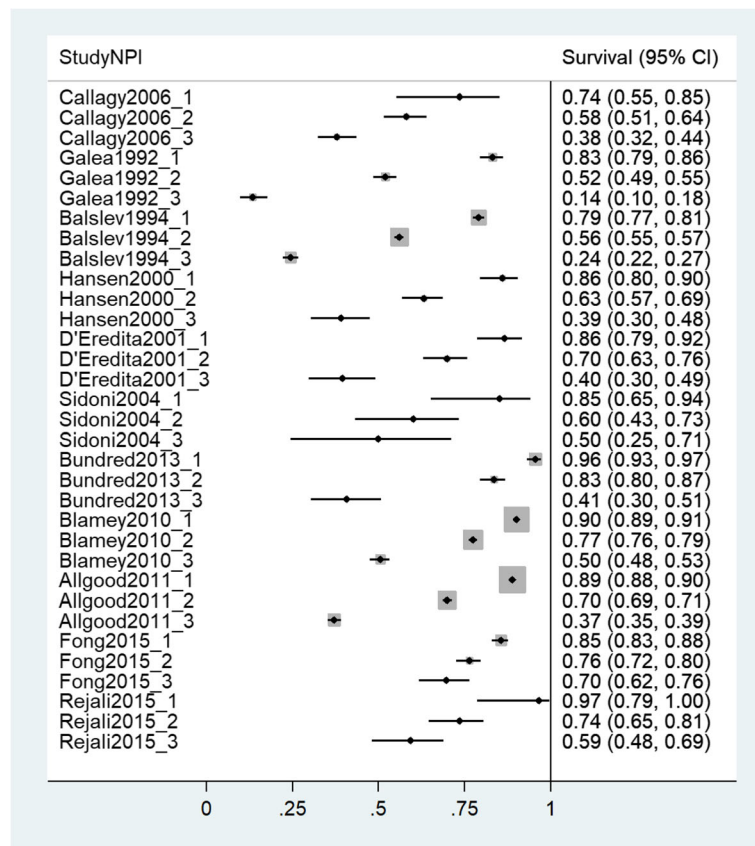


Fig. 4 Forest plot of 10-year survival estimates from individual studies. Studies presented in order of mid-point of data collection earliest to most recent. X-axis shows proportions of sample surviving at the specified time point

some of these limitations. The resulting exclusions may introduce some selection bias, but it is not possible to know the impact of such a bias.

There were some observed differences in included studies in the methods used to assess lymph node status

which is a key input to define NPI category. Originally, node involvement for selected NPI category was assessed using biopsy results of a lower axillary node, an apical axillary node and a node from the internal mammary chain. Node status was specified as one if the tumour

Table 2 Results from the meta-analysis, pooled survival estimates and heterogeneity statistics

Statistic	NPI 1	NPI 2	NPI 3
5-year follow-up			
Fixed effects I-V estimate (95% CI)	0.953 (0.949, 0.957)	0.831 (0.825, 0.837)	0.535 (0.521, 0.549)
Random effects D-L estimate (95% CI)	0.943 (0.927, 0.96)	0.811 (0.769, 0.853)	0.565 (0.462, 0.668)
I^2 (%)	89.6	97.2	97.7
Q	115.93 ($P < 0.001$)	422.5 ($P < 0.001$)	532.32 ($P < 0.001$)
Tau (% points)	2.45	7.28	18.25
10-year follow-up			
Fixed effects I-V estimate (95% CI)	0.883 (0.877, 0.888)	0.682 (0.675, 0.689)	0.353 (0.341, 0.364)
Random effects D-L estimate (95% CI)	0.869 (0.837, 0.901)	0.674 (0.608, 0.74)	0.414 (0.329, 0.499)
I^2 (%)	95.3	98.6	97.7
Q	213.56 ($P < 0.001$)	727.14 ($P < 0.001$)	427.44 ($P < 0.001$)
Tau (% points)	4.58	10.77	13.71

I-V inverse variance, D-L DerSimonian and Laird method

was absent from all three nodes, two if tumour cells were found in lower axillary node only, and three if tumour cells were in the apical and/or internal mammary node [4]. To increase usability of the NPI, Galea et al. [5] suggested staging based on the number of involved nodes from the single location as a viable alternative, which was subsequently used in later studies. However, we believe this change in assessment of node status was sufficiently similar to not introduce bias in the analysis.

Conclusions

Precision medicine promises to deliver improved patient outcomes through better understanding of patient and disease heterogeneity, and ultimately better targeted screening and therapeutic strategies. Moving the concept of precision medicine into clinical practice requires mechanisms to identify and stratify eligible patient populations. Prognostic models are an important component in such a mechanism because they allow an individualized quantitative estimate of potential treatment benefit.

This systematic review and meta-analysis of NPI stratified estimates of survival revealed a key challenge in the design and evaluation of precision medicine interventions; that substantial differences may exist between studies estimating the relationship between a marker or score and the outcome of interest. Heterogeneity of study estimates should be investigated carefully for all such interventions. This finding suggests a potential additional burden on shared decision-making between clinicians and patients. Achieving a better informed decision is only possible if all of the available evidence is correctly synthesized and clearly presented.

The results of this systematic review and meta-analysis have implications for patients, clinicians, decision analysts and policy analysts requiring a prediction of long-term overall survival, in women diagnosed with early and/or locally advanced breast cancer. The observed heterogeneity in overall survival estimates indicates that it is important to make use of survival data from an appropriate setting to provide reliable prognostic information for use in clinical practice, and also in decision-analytic model-based economic evaluations of new treatments, screening programmes or prevention strategies to inform health care resource allocation decisions.

Endnotes

¹A study that takes a group (cohort) of patients and follows their progress over time in order to measure outcomes such as disease or mortality rates and make comparisons according to the treatments or interventions that patients received [47].

²Cancer reappearing after a remission [48].

³A proliferation of malignant-appearing cells of the ducts and terminal lobular units of the breast that have not breached the ductal basement membrane [49].

Additional files

Additional file 1: Appendix 1. NPI category cut-offs. **Appendix 2.** Electronic search strategies. **Appendix 3.** Critical appraisal of accepted studies. **Appendix 4.** Data extraction procedures. **Appendix 5.** Reported five and ten-year survival per study dataset. (DOCX 71 kb)

Additional file 2: PRISMA checklist. (DOCX 21 kb)

Abbreviations

CASP: Critical Skills Appraisal Programme; HER-2: Human epidermal growth factor receptor 2; NPI: Nottingham Prognostic Index; PROSPERO: International prospective register of systematic reviews; RCT: Randomized controlled trial; TNM: Tumour, nodes, metastases

Funding

The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7 under grant agreement no. 306088. The funder had no role in the design of the study or collection, analysis, and interpretation of data, or in writing the manuscript.

Availability of data and materials

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

EG contributed to the design of the study, abstract and full text review; design of the statistical analysis; and writing of the manuscript. AD contributed to the literature search, abstract and full text review; data extraction; data analysis; and writing of the manuscript. PH contributed to interpretation of the results and writing of the manuscript. KP contributed to the design of the study, interpretation of the results and writing of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹The University of Edinburgh, Edinburgh, UK. ²Manchester Centre for Health Economics, The University of Manchester, Manchester, UK.

Received: 16 August 2017 Accepted: 28 August 2018

Published online: 15 September 2018

References

- Blamey RW, Hornmark-Stenstam B, Ball G, Blichert-Toft M, Cataliotti L, Fourquet A, et al. ONCOPOOL - a European database for 16,944 cases of breast cancer. *Eur J Cancer Elsevier Ltd.* 2010;46:56–71.
- Lundin J, Lehtimäki T, Lundin M, Holli K, Elomaa L, Turpeenniemi-Hujanen T, et al. Generalisability of survival estimates for patients with breast cancer--a comparison across two population-based series. *Eur J Cancer.* 2006;42:3228–35.
- Sobin L, Gospodarowicz M, Wittekind C. TNM classification of malignant tumours. 2011.

4. Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. *Br J Cancer*. 1982;45:361–6.
5. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat*. 1992;22:207–19.
6. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res*. 2010;12:R1.
7. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19:980–91.
8. Rakha EA, Soria D, Green a R, Lemetre C, Powe DG, Nolan CC, et al. Nottingham Prognostic Index Plus (NPI+): a modern clinical decision making tool in breast cancer. *Br J Cancer*. 2014;110:1688–97. Nature Publishing Group
9. National Institute for Health and Care Excellence. CG80 early and locally advanced breast cancer: diagnosis and treatment. 2009;
10. The Precision Medicine Initiative. Precision Medicine Initiative | The White House. 2016. Available from: <https://obamawhitehouse.archives.gov/precision-medicine>. [cited 8 May 2017]
11. Blamey RW. 4. The design and clinical use of the Nottingham Prognostic Index in breast cancer. *Breast*. 1996;5:156–7. Elsevier
12. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal, Process and methods guides, vol. 9; 2013.
13. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ*. 2006;15:677–87.
14. University of York. NHS Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2009.
15. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *J Am Med Assoc*. 2000;283:2008–12.
16. Nelson H, Tyne K, Naik A, Bougatso C, Chan B, Nygren P, et al. Screening for breast cancer: systematic evidence review update for the U.S. Preventive Services Task Force. Evidence review update no. 74. 2009.
17. The CASP Team. Critical Appraisal Skills Programme (CASP) Cohort Study Checklist. 2013. Available from: <https://casp-uk.net/casp-tools-checklists/>. [cited 2 Feb 2017]
18. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12:9.
19. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–88.
20. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated march 2011]. The Cochrane Collaboration; 2011.
21. Borenstein M, Higgins JPT, Hedges LV, Rothstein HR. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res Synth Methods*. 2017;8:5–18.
22. Gavaghan DJ, Moore RA, McQuay HJ. An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain*. 2000;85:415–24.
23. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539–58.
24. Rice K, Higgins JPT, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2018;181:205–27.
25. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health care evaluation. *Stat. Pract*. 2004.
26. Blamey RW, Ellis IO, Pinder SE, Lee AHS, Macmillan RD, Morgan DAL, et al. Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990–1999. *Eur J Cancer*. 2007;43:1548–55.
27. Suen D, Chow LWC. Prognostic contribution of the HER-2 oncogene overexpression to the Nottingham Prognostic Index in breast cancer. *Biomed Pharmacother*. 2006;60:293–7.
28. Jarman IH, Etechells TA, Martín JD, Lisboa PJG. An integrated framework for risk profiling of breast cancer patients following surgery. *Artif Intell Med*. 2008;42:165–88.
29. Wishart GC, Greenberg DC, Britton PD, Chou P, Brown CH, Purushotham AD, et al. Screen-detected vs symptomatic breast cancer: is improved survival due to stage migration alone? *Br J Cancer*. 2008;98:1741–4.
30. Anderson TJ, Alexander FE, Lamb J, Smith A, Forrest AP. Pathology characteristics that optimize outcome prediction of a breast screening trial. *Br J Cancer*. 2000;83:487–92.
31. Sauerbrei W, Hübner K, Schmoor C, Schumacher M. Validation of existing and development of new prognostic classification schemes in node negative breast cancer. *Breast Cancer Res Treat*. 1997;42:149–63.
32. Hwang KT, Woo JW, Shin HC, Kim HS, Ahn SK, Moon HG, et al. Prognostic influence of BCL2 expression in breast cancer. *Int J Cancer*. 2012;131:E1109–19.
33. Campbell H, Gray A, Harris A, Briggs A, Taylor M. Estimation and external validation of a new prognostic model for predicting recurrence-free survival for early breast cancer patients in the UK. *Br. J. Cancer*. 2010;103:776–86. Nature Publishing Group
34. Sundquist M, Thorstenson S, Brudin L, Nordenskjöld B, Group SESBCS. Applying the Nottingham Prognostic Index to a Swedish breast cancer population. *Breast Cancer Res Treat*. 1999;53:1–8.
35. Hasebe T, Sasaki S, Imoto S, Wada N, Ishii G, Ochiai A. Primary tumour–vessel tumour–nodal tumour classification for patients with invasive ductal carcinoma of the breast. *Br J Cancer*. 2005;92:847–56.
36. Quintyne KI, Woulfe B, Coffey JC, Gupta RK. Correlation between Nottingham Prognostic Index and Adjuvant! Online prognostic tools in patients with early-stage breast cancer in Mid-Western Ireland. *Clin Breast Cancer*. 2013;13:233–8. Elsevier
37. Winzer KJ, Buchholz A, Schumacher M, Sauerbrei W. Improving the prognostic ability through better use of standard clinical data - the Nottingham Prognostic Index as an example. *PLoS One*. 2016;11:e0149977.
38. Kollias J, Elston CW, Ellis IO, Robertson JF, Blamey RW. Early-onset breast cancer—histopathological and prognostic considerations. *Br J Cancer*. 1997; 75:1318–23.
39. Kollias J, Murphy CA, Elston CW, Ellis IO, Robertson JFR, Blamey RW, et al. The prognosis of small primary breast cancers. *Eur J Cancer*. 1999;35:908–12.
40. Blamey RW, Pinder SE, Ball GR, Ellis IO, Elston CW, Mitchell MJ, et al. Reading the prognosis of the individual with breast cancer. *Eur J Cancer*. 2007;43: 1545–7.
41. Allgood PC, Duffy SW, Kearins O, O'Sullivan E, Tappenden N, Wallis MG, et al. Explaining the difference in prognosis between screen-detected and symptomatic breast cancers. *Br J Cancer*. 2011;104:1680–5.
42. Callagy GM, Pharoah PD, Pinder SE, Hsu FD, Nielsen TO, Ragaz J, et al. Bcl-2 is a prognostic marker in breast cancer independently of the Nottingham Prognostic Index. *Clin Cancer Res*. 2006;12:2468–75.
43. Rejali M, Tazhibi M, Mokarian F, Gharanjik N, Mokarian R. The performance of the Nottingham Prognosis Index and the adjuvant online decision making tool for prognosis in early-stage breast cancer patients. *Int J Prev Med*. 2015;6:93.
44. Bundred NJ, Prasad R, Morris J, Knox WF, Byrne G, Cheung S, et al. Are symptomatic guidelines for chemotherapy appropriate to ER-positive screen-detected breast cancer (SDBC)? *Breast Cancer Res Treat*. 2013;138: 359–68.
45. Hansen S, Grabau DA, Sørensen FB, Bak M, Vach W, Rose C. Vascular grading of angiogenesis: prognostic significance in breast cancer. *Br J Cancer*. 2000; 82:339–47.
46. Fong Y, Evans J, Brook D, Kenkre J, Jarvis P, Gower-Thomas K. The Nottingham Prognostic Index: five- and ten-year data for all-cause survival within a screened population. *Ann R Coll Surg Engl*. 2015;97:137–9.
47. National Institute for Health and Care Excellence. Glossary | NICE. NICE; Available from: <https://www.nice.org.uk/glossary>. [cited 2 Feb 2017]
48. Stedman TL. Stedman's medical dictionary. Philadelphia: Lippincott Williams & Wilkins; 2006.
49. Olivetto I, Levine M, Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer. Clinical practice guidelines for the care and treatment of breast cancer: the management of ductal carcinoma in situ (summary of the 2001 update). *CMAJ*. 2001;165:912–3.