European Journal of Public Health, Vol. 34, No. 1, 44–51

https://doi.org/10.1093/eurpub/ckad181 Advance Access published on 24 October 2023

European value-based healthcare benchmarking: moving from theory to practice

Borja García-Lorenzo [1], Ania Gorostiza[1], Itxaso Alayo[1], Susana Castelo Zas[2], Patricia Cobos Baena[3], Inés Gallego Camiña[2], Begoña Izaguirre Narbaiza[4], Gaizka Mallabiabarrena[3], Iker Ustarroz-Aguirre[5], Alina Rigabert[6], William Balzi[7], Roberta Maltoni[7], Ilaria Massa[7], Isabel Álvarez López[8,9], Sara Arévalo Lobera[8,9], Mónica Esteban[10], Marta Fernández Calleja[11], Jenifer Gómez Mediavilla[8], Manuela Fernández[10], Manuel del Oro Hitar[12,13], María del Carmen Ortega Torres[12,13], María Consuelo Sanz Ferrandez[12], Luís Manso Sánchez[13,14], Pablo Serrano Balazote[13,14], Carolina Varela Rodríguez[13,14], Mario Campone [15], Sophie Le Lann[15], Piet Vercauter[16], Kurt Tournoy[16,17], Marina Borges[18], Ana Sofía Oliveira[18], Marta Soares[18], Ane Fullaondo[1], on behalf of the VOICE Study Group*

1 Biosistemak Institute for Health Systems Research, Torre del Bilbao Exhibition Centre, Barakaldo, Spain
2 Osakidetza Basque Health Service, Ezkerraldea Enkarterri Cruces Integrated Health Organisation, Innovation and Quality Assistant Service, Barakaldo-Bizkaia, Spain
3 Osakidetza Basque Health Service, Ezkerraldea Enkarterri Cruces Integrated Health Organisation, Mammary Pathology Service, Barakaldo-Bizkaia, Spain
4 Osakidetza Basque Health Service, Ezkerraldea Enkarterri Cruces Integrated Health Organisation, Analytical Accounting, Economic and Financial Directorate, Barakaldo, Spain
5 Osakidetza Basque Health Service, Ezkerraldea Enkarterri Cruces Integrated Health Organisation, Economic Evaluation Unit, Economic and Financial Directorate, Barakaldo, Spain
6 Fundación Andaluza Beturia para la Investigación en Salud (FABIS), Huelva, Spain
7 IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
8 Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Medical Oncology, Donostia, Spain
9 Biodonostia, Donostia, Gipuzkoa, Spain
10 Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Economic Resource Service, Donostia, Gipuzkoa, Spain
11 Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Mamary Pathology Service, Donostia, Gipuzkoa, Spain
12 Gynecology Department, Hospital Universitario 12 de Octubre, Madrid, Spain
13 Hospital Universitario 12 de Octubre, Madrid, Spain
14 Instituto de Investigación Biomédica del Hospital Universitario 12 de Octubre, Madrid, Spain
15 Institut de Cancérologie de l'Ouest, Angers-Nantes, France
16 Department of Pulmonary Medicine, Onze-Lieve-Vrouw Hospital, Aalst, Belgium
17 Faculty of Medicine and Life Sciences, Ghent University, Ghent, Belgium
18 Instituto Português de Oncologia do Porto, Portugal

Correspondence: Borja García-Lorenzo, Biosistemak Health Systems Research Institute, Azkue Kalea, 1, Torre del Bilbao Exhibition Centre, 48902 Barakaldo, Basque Country, Spain, Tel: +34 63939490, e-mail: bgarcia@kronikgune.org
*The members of the VOICE Study Group are listed in the Acknowledgements.

Background: Value-based healthcare (VBHC) is a conceptual framework to improve the value of healthcare by health, care-process and economic outcomes. Benchmarking should provide useful information to identify best practices and therefore a good instrument to improve quality across healthcare organizations. This paper aims to provide a proof-of-concept of the feasibility of an international VBHC benchmarking in breast cancer, with the ultimate aim of being used to share best practices with a data-driven approach among healthcare organizations from different health systems. **Methods:** In the VOICE community—a European healthcare centre cluster intending to address VBHC from theory to practice—information on patient-reported, clinical-related, care-process-related and economic-related outcomes were collected. Patient archetypes were identified using clustering techniques and an indicator set following a modified Delphi was defined. Benchmarking was performed using regression models controlling for patient archetypes and socio-demographic characteristics. **Results:** Six hundred and ninety patients from six healthcare centres were included. A set of 50 health, care-process and economic indicators was distilled for benchmarking. Statistically significant differences across sites have been found in most health outcomes, half of the care-process indicators, and all economic indicators, allowing for identifying the best and worst performers. **Conclusions:** To the best of our knowledge, this is the first international experience providing evidence to be used with VBHC benchmarking intention. Differences in indicators across healthcare centres should be used to identify best practices and improve healthcare quality following further research. Applied methods might help to move forward with VBHC benchmarking in other medical conditions.

Introduction

Cancer policy frameworks underline that cancer care is not exempt from the problem of the overuse or underuse of procedures and interventions. Policymakers strive to measure and improve the quality of cancer care systematically. This attempt to reallocate healthcare resources towards efficient and appropriate healthcare entails the awareness of value-based healthcare (VBHC) on cancer.[1] VBHC puts patient outcomes at the centre of healthcare, linking outcomes to costs and therefore determining the value of the healthcare. The key concept behind VBHC is to improve achieved outcomes *per* encountered costs and evaluate the performance accordingly.[2] Being aware of the VBHC on cancer might allow benchmarking across healthcare centres and, consequently, identify lessons that may be learned from the best performers.[3] To reach this point, the accurate and systematic measurement of health and economic outcomes turns up as an essential requirement.[2,4,5] The standardization and systematic recording of patient outcomes allow exploring the patients' perception of their quality of life to redirect the healthcare towards a shared decision-making process. The aforementioned standardization might ultimately allow healthcare through outcomes-based payments.

Even though the survival rate has considerably increased in breast cancer over recent years,[6] a significant decrease in the quality of life[7] remains, together with high healthcare resource utilization.[8] Health outcomes, defined as patient-reported outcomes (PROs) and clinical-related outcomes (CROs), have been standardized by the International Consortium for Health Outcomes Measurement (ICHOM) to become a solid basis for comparison.[9] Care-process-related outcomes (CAPROs) have been proposed by the European Society of Breast Cancer Specialists (EUSOMA).[10] Despite the development of the above-mentioned standardized tools for outcome measures, there is a lack of VBHC benchmarking in breast cancer. Although the use of clinical practice guidelines, there exists variability in clinical practice.[11] Benchmarking is a valid instrument for detecting these differences, attributing processes to results and improving, often by learning from others. Collaboration, feedback and reflective practice within networks is imperative to achieve a healthcare system that is less prone to unwarranted variation. Benchmarking should provide useful information to identify best practices, and therefore a good instrument to improve quality across healthcare organizations.[3] Particularly, VBHC benchmarking in breast cancer has become a challenge following a threefold reason: (i) the low adherence to systematically measuring patient health outcomes by healthcare centres, (ii) the lack of a standard set of healthcare economic-related outcomes (EROs) and (iii) the absence of a validated set of breast cancer indicators including health, care-process and economic outcomes. This paper aims to address the aforementioned lack by providing the proof-of-concept of the feasibility of an international VBHC benchmarking of health, care-process and economic outcomes in breast cancer. This purpose has been developed into the VOICE community (Value-Based Healthcare for Outcomes in Breast and Lung Cancer in Europe), an eight-European healthcare centre cluster that intends to address the VBHC in breast and lung cancer from theory into practice.

Methods

Study design

This study was designed as a prospective multicentre cohort across six pilot sites in three countries: *Organización Sanitaria Integrada Ezkerraldea-Enkarterri-Cruces*, *Organización Sanitaria Integrada Donostialdea*, *Hospital Juan Ramón Jiménez* and *Hospital Universitario 12 de Octubre* (Spain), *Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS): Istituto Romagnolo per lo Studio dei Tumori (IRST) 'Dino Amadori'* (Italy) and *Institut de Cancérologie de l'Ouest* (France), with the approval of the Basque Country Ethics Committee (PI2018107).

The study population consists of patients diagnosed with early breast cancer between 2018 and 2020, meeting the following eligibility criteria: (i) aged over 18 years, (ii) female, (iii) newly diagnosed invasive breast cancer (stage I–III) or ductal carcinoma *in situ* (DCIS) and (iv) with any treatment (surgery, radiotherapy, chemotherapy, hormone therapy and/or targeted therapy). Criteria for exclusion included patients with (i) rare tumours, (ii) lobular carcinoma *in situ* or (iii) metastatic disease. Candidates were progressively recruited and contacted at medical/nursing visits by their healthcare professionals at diagnosis and provided with the VOICE community information. Patients were recruited at diagnosis and followed for 6 months after diagnosis or until death, whichever came first. Participants signed informed consent at recruitment.

Data collection

Assessment was based on analyzing data from three sources: the questionnaires completed by the patients, the electronic health records and the administrative databases.

Socio-demographic variables, PROs, CROs and CAPROs were collected according to data collection guide of the ICHOM standard set[9,12]: socio-demographic variables were collected at baseline, defined as the diagnosis date. PROs, CROs and CAPROs information were collected at baseline and followed-up at 6 months. The ICHOM set, a minimal standard set of PROs, CROs and CAPROs deemed to be most appropriate to patients with breast cancer, and generally applicable worldwide, was designed through the use of a literature review, an extensive patient input and the consensus of a multidisciplinary international working group comprised 26 healthcare providers and patient advocates. The working group convened via eight teleconferences and followed a modified two-round Delphi method to achieve consensus.

EROs, resource use and unit costs were collected at a 12-month follow-up. Since no standard set on economic outcomes was available for breast cancer, an *ad hoc* set of resource use was designed following the patient care pathway designed by the VOICE research team,[13,14] the healthcare centre perspective and the health economics literature on healthcare utilization.[8,15] A reference group from the VOICE community involving four health economists and four clinicians, generated a draft list of resource use and their corresponding unit costs to explore, on the one hand, the availability of the economic information and, on the other, the feasibility of a *bottom-up* approach[16] to estimate the cost per patient. To deal with the heterogeneous availability in the detail level of resource use, a minimum database was defined by the reference group to allow for comparability. Supplementary table S1 shows the codebook of EROs collected.

Patient data used in the analyses were anonymized. In order to ensure that there was no intention to re-identify individuals, Data Processing Agreements were bilaterally formalized between the data processor and each data provider.

Benchmarking indicators set

Following a Delphi methodology,[17,18] a two-step process was undertaken to choose an appropriate, manageable and relevant indicator set based on the collected outcomes to conduct a benchmarking moving towards the promotion of patient-centred outcomes (PCOs) in the context of the VBHC.

PCOs are being integrated into policy initiatives at the national or international levels in many industrialized countries.[19] Previous to the modified Delphi, the initial set of variables (323) was transformed into a preliminary set of health, care-process and economic indicators (78) as follows: The health (PROs and CROs) indicators were directly derived from the standard set for breast cancer defined by ICHOM.[9] PROs were derived from the EORTC QLQ-C30/BR23/

LMC21, BREAST-Q and FACT-6 questionnaire and transformed into health indicators computing the patient score proposed by their corresponding questionnaires and computing the patient average by each healthcare organization. CROs were transformed into health indicators following the ratio proposed by ICHOM.[9] The care-process indicators were selected from the EUSOMA guideline, which proposes robust and evidence-based quality indicators in breast cancer care under their computation feasibility based on the dataset collected according to the ICHOM standard set. CAPROs were transformed into care-process indicators following the proposed indicators of the clinical practice guideline[10] of the EUSOMA. Finally, the economic indicators were based on the EROs designed by the VOICE research team following the health economics literature on healthcare utilization.[8,15] EROs were transformed into economic indicators computing the patient average or ratio by each healthcare organization. Supplementary figure S1 and Tables S1 and Table S2 show the patient pathway, the EROs and the list of the proposed indicators, respectively.

Step 1

A preliminary list of 78 indicators containing health, care-process and economic indicators was proposed to the VOICE research group, formed by seven clinicians (gynaecologists, medical oncologists, internal physicians and radio-oncologists), two healthcare managers, two health economists and one statistician. They were selected following a purposeful sampling, with the aim to identify and select individuals who are especially aware or informed about or experienced breast cancer management.

Information on the preliminary proposed indicators was previously and individually distributed to the VOICE research group as the basis for discussion at the meeting. Experts were asked to provide their opinions on each indicator following group dynamics which allowed participants to pool their expertise and open discussions to make a decision on the corresponding indicators, with the aim of designing a synthetic and relevant indicator set based on their contribution to the healthcare quality in the European context, excluding indicators being merely an overall basic standard compliance. The discussion allowed for the outright rejection of some indicators and the definition of an indicator set proposal on which to base the benchmarking.

Step 2

The second step involved an expert panel to rate every agreed indicator in Step 1. This panel of experts was integrated by 20 clinicians and 5 healthcare managers, with various perspectives and expertise, to produce unbiased consensual knowledge. They were selected following a purposeful sampling across pilot sites, with the aim to identify and select individuals who are especially aware, informed and experienced breast cancer management. A glossary of indicators was provided to the expert panel to understand the concepts used in the questionnaire and reduce the risk of bias due to semantic issues. All expert panel members were contacted via email, the purpose of the research was explained and their consent was obtained.

The expert panel was asked to rate each indicator on a five-point Likert scale[20] with the following response options: strongly agree, agree, neutral, disagree and strongly disagree. The level of agreement had to address the relevance of the indicator to be considered in a VBHC benchmarking across healthcare centres. The experts were also asked to express their comments on each indicator if convenient. Consensus to include the indicator was defined as when at least 70%[16] of experts select the strongly agree or agree categories. The indicators with no consensus were discussed by the VOICE study group and finally excluded. Supplementary table S2 shows the preliminary list of indicators and the indicators included by consensus.

Benchmarking analysis

Patient archetypes

Archetypes based on patients' characteristics and their healthcare pathway were defined to ensure health outcomes' comparability between sites. Current patient classifications are mainly based on biological and molecular subtypes; however, there is no patient classification so far which directly allows us to relate the patient journey, which is considerably influenced by the therapeutic actions, and the patients perceived outcomes. Therefore, the rationale behind these patient archetypes lies in the need to consider the factors that primarily affect patients' overall perceived quality of life, the cornerstone of VBHC.

A mixed approach combining clinical criteria and data mining techniques was used to obtain clinically validated patient archetypes. Patients presenting only ductal carcinoma *in situ* (DCIS) were grouped following clinical experts' criteria. Clinicians selected a set of variables from the dataset for archetypes definition of patients with invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC) or other carcinomas (OC). Demographic (*age*) and treatment-related (*type of surgery*, *type of axillary surgery*, *chemotherapy*, *radiotherapy*, *hormonotherapy*) variables were selected. IDC, ILC, and OC patients were clustered based on their similarity regarding previous variables following a Hierarchical Clustering on Principal Components (HCPC) technique.[21,22] Archetypes obtained from HCPC were discussed and validated by the clinicians of the VOICE research team, who decided to split one of the archetypes into two to ensure their clinical significance. See 'Patient archetypes' in 'Results' section.

Descriptive analysis

A descriptive analysis of clinical and socio-demographic variables across sites was carried out, followed by chi-square statistical tests to assess their statistical differences.

Regression analysis

Benchmarking was based on the comparison across sites of previously selected indicators. To account for patient heterogeneity and avoid potentially misleading conclusions, comparisons were risk-adjusted.[23–25] A multivariate regression analysis indicator-by-indicator[26–29] was conducted to explore differences in indicators across sites, controlling for patient heterogeneity through archetypes and clinical, and socio-demographic patient characteristics. In the case of PROs, each indicator was also adjusted for its value at baseline. Indicators are typically calculated and disseminated at the hospital level. However, at patient level, the outcome can be either a numerical or a dichotomous variable. According to the indicator measurement—continuous, binary or count—linear, logistic or Poisson regression models were correspondingly estimated by ordinary least squares or maximum likelihood in Eqs (1), (2) and (3) as follows:

$$I_i = \alpha_0 + \sum_{k=1}^{K} \beta_k Site_{ki} + \sum_{h=1}^{H} \delta_H Archetype_{hi} + \sum_{w=1}^{W} \varnothing_w X_{wi} + \mu_i \quad \text{Eq. (1)}$$

$$\begin{aligned} \mathrm{Ln}\left(\frac{P(I_i = 1)}{1 - P(I_i = 1)}\right) &= \alpha_0 + \sum_{k=1}^{K} \beta_k Site_{ki} \\ &+ \sum_{h=1}^{H} \delta_H Archetype_{hi} + \sum_{w=1}^{W} \varnothing_w X_{wi} + \mu_i \end{aligned} \quad \text{Eq. (2)}$$

$$\mathrm{Ln}(P(I_i = z)) = \alpha_0 + \sum_{k=1}^{K} \beta_k Site_{ki} + \sum_{h=1}^{H} \delta_H Archetype_{hi} + \sum_{w=1}^{W} \varnothing_w X_{wi} + \mu_i \quad \text{Eq. (3)}$$

where I_i is defined as the indicator value for patient i, $P(I_i = 1)$ the probability of I_i being equal to 1, $P(I_i = z)$ the probability of I_i being equal to z, and Ln, the Napierian logarithm operator. $P(I_i = 1)$ follows a logistic distribution while $P(I_i = z)$ so does a Poisson distribution assuming no overdispersion.[30] Covariate $Site_i$ is defined as a categorical variable identifying the healthcare centre, $Archetype_i$ is defined as a categorical variable identifying the defined archetypes (see 'Patient archetypes' in 'Results' section for further information), and X_i is defined as the vector of the control variables *age*, *menopause*, *comorbidities*, *laterality*, *grade of invasive component, oestrogen, progesterone*, and *HER2 receptor* status, defined in the data collection reference guide of ICHOM.[12] α_0, β_K, δ_H and ϕ_W are the constant term and the parameters associated with covariates $Site_i$, $Archetype_i$ and X_i, respectively, while μ_i is the random error term. Quantitative and statistical differences across the β_K parameters for each site are intended to illustrate the benchmarking. Either a single site or a group of sites were identified as best/worst performers' indicator-by-indicator.

For each indicator, a full model including all potential covariates was first estimated, then a reduced model was estimated using a stepwise regression strategy. Regression-based test was conducted to test for overdispersion in Poisson regression models.[31] Negative binomial regression model, where $P(I_i = z)$ in Eq. (3) follows a Pascal distribution, was used to account for overdispersion when identified. The statistical significance was set at P values <0.05. Analyses were performed using R version 4.2.2.

Results

Benchmarking indicator set

The *Step 1* analysis revealed a set proposal of 54 indicators. In *Step 2*, consensus was reached in the first round for 50 indicators. The remaining four indicators with no consensus were care-process indicators and were not finally included following the aim of reaching a manageable indicator set and the promotion of PCOs in the VBHC context. Thus, a set of 50 indicators classified in health outcomes (25), care-process (8), economic (11) and risk-adjustment (6) was selected.

Benchmarking analysis

Patient archetypes

Patients were classified into eight clinically validated archetypes. DCIS patients were grouped in archetype *Arch-1* following the clinical experts' criteria. The HCPC technique identified six clinically relevant profiles for IDC, ILC and OC patients. Arch-3 was split into *Arch-3.1* and *Arch-3.2* based on clinical criteria, with the *Arch-3.1* patients undergoing chemotherapy treatment while *Arch-3.2* not. Five patients were not assigned to any archetype due to missing data. Patient archetypes and their corresponding most representative patient in each archetype are described in table 1.

Descriptive analysis

Six hundred and ninety patients were recruited. The average age was 59 (SD = 12); 42.3% reported comorbidities and 66% had postmenopausal status. Statistically significant heterogeneity across sites was found in all patient characteristics except for *Menopause status*, *Positive oestrogen receptor status* and *chemotherapy*. The sample description is shown in table 2. Six hundred and thirty-eight (92.5%) completed all the PROs at baseline, and 616 (96.6%) completed at least one PROs in a 6-month follow-up. Supplementary tables S3 and S4 show the statistical description of the PROs at baseline and 6 months, and the rest of health, care-process and the economic indicators, respectively.

Regression analysis

Statistically significant differences across sites have been found in most health outcomes, half of care-process indicators and all economic indicators.

Regarding the health indicators, all PROs have shown significant differences across sites except *sexual functioning* and *breast satisfaction*. One single best/worst performer was found in *physical functioning* (e.g. *site D* scoring on average 11 points lower than the rest of the sites), *body image*, *insomnia, vaginal symptoms* and *vasomotor symptoms.* A group of best/worst performers has been identified in *overall well-being* (e.g. *site D* and *site E* scoring lower than the rest of sites), *emotional*, *cognitive,* and *social functioning*, *financial impact*, *fatigue*, *pain*, *arthralgia*, and *breast*, *arm,* and *peripheral symptoms*. The only

Table 1 Patient profiles across archetypes[a]

	Arch-1 $N_1 = 102$ **(14.9%)**	**Arch-2** $N_2 = 84$ **(12.3%)**	**Arch-3.1** $N_{3.1} = 68$ **(9.9%)**	**Arch-3.2** $N_{3.2} = 198$ **(28.9%)**	**Arch-4** $N_4 = 54$ **(7.9%)**	**Arch-5** $N_5 = 58$ **(8.5%)**	**Arch-6** $N_6 = 51$ **(7.4%)**	**Arch-7** $N_7 = 70$ **(10.2%)**
Age at diagnosis: (*N*=685)								
≤50	21 (20.6%)	0 (0.00%)	32 (47.1%)	46 (23.2%)	13 (24.1%)	21 (36.2%)	27 (52.9%)	34 (48.6%)
51–70	63 (61.8%)	0 (0.00%)	36 (52.9%)	148 (74.7%)	27 (50.0%)	18 (31.0%)	20 (39.2%)	35 (50.0%)
>70	18 (17.6%)	84 (100%)	0 (0.00%)	4 (2.02%)	14 (25.9%)	19 (32.8%)	4 (7.84%)	1 (1.43%)
Ductal carcinoma in situ (*N*=685)	102 (100%)	3 (3.57%)	6 (8.82%)	11 (5.56%)	2 (3.70%)	4 (6.90%)	6 (11.8%)	9 (12.9%)
Invasive ductal carcinoma (*N*=685)	0 (0.00%)	63 (75.0%)	64 (94.1%)	164 (82.8%)	49 (90.7%)	47 (81.0%)	43 (84.3%)	61 (87.1%)
Invasive lobular carcinoma (*N*=684)	0 (0.00%)	17 (20.2%)	2 (2.94%)	23 (11.6%)	1 (1.85%)	8 (13.8%)	5 (9.80%)	10 (14.3%)
Other carcinoma (*N*=685)	0 (0.00%)	4 (4.76%)	3 (4.41%)	12 (6.06%)	4 (7.41%)	3 (5.17%)	5 (9.80%)	1 (1.43%)
Surgery (*N*=685)								
BCS	75 (73.5%)	81 (96.4%)	60 (88.2%)	189 (95.5%)	45 (83.3%)	0 (0.00%)	1 (1.96%)	29 (41.4%)
BCS with mammoplasty	4 (3.92%)	3 (3.57%)	8 (11.8%)	8 (4.04%)	6 (11.1%)	0 (0.00%)	0 (0.00%)	4 (5.71%)
Mastectomy without immediate reconstruction	9 (8.82%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	58 (100%)	0 (0.00%)	0 (0.00%)
Mastectomy with immediate reconstruction	14 (13.7%)	0 (0.00%)	0 (0.00%)	1 (0.51%)	3 (5.56%)	0 (0.00%)	50 (98.0%)	37 (52.9%)
Surgery to axilla (*N*=683)								
None	54 (53.5%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3 (5.17%)	0 (0.00%)	0 (0.00%)
Sentinel lymph node biopsy	29 (28.7%)	40 (48.2%)	40 (58.8%)	146 (73.7%)	23 (42.6%)	14 (24.1%)	36 (70.6%)	6 (8.57%)
Axillary sampling	15 (14.9%)	32 (38.6%)	28 (41.2%)	44 (22.2%)	17 (31.5%)	12 (20.7%)	8 (15.7%)	4 (5.71%)
Axillary clearance	3 (2.97%)	11 (13.3%)	0 (0.00%)	8 (4.04%)	14 (25.9%)	29 (50.0%)	7 (13.7%)	60 (85.7%)
Radiotherapy (*N*=685)	75 (73.5%)	73 (86.9%)	67 (98.5%)	194 (98.0%)	49 (90.7%)	35 (60.3%)	0 (0.00%)	70 (100%)
Chemotherapy (*N*=685)	3 (2.94%)	11 (13.1%)	68 (100%)	0 (0.00%)	44 (81.5%)	36 (62.1%)	16 (31.4%)	66 (94.3%)
Hormonotherapy (*N*=685)	46 (45.1%)	84 (100%)	68 (100%)	198 (100%)	2 (3.70%)	48 (82.8%)	47 (92.2%)	65 (92.9%)

Note: BCS, breast conserving surgery.

a: Percentages calculated excluding missing data.

Table 2 Descriptive analysis[a]

	Total N=690	Site A N_A =273 (39.6%)	Site B N_B = 58 (8.4%)	Site C N_C = 42 (6.1%)	Site D N_D = 99 (14.3%)	Site E N_E = 127 (18.4%)	Site F N_F = 91 (13.2%)	*P*-value[b]
Patient characteristics								
Age at diagnosis (*N*=690)								<0.001***
≤50	198 (28.7%)	80 (29.3%)	23 (39.7%)	16 (38.1%)	33 (33.3%)	26 (20.5%)	20 (22.0%)	
51–70	348 (50.4%)	152 (55.7%)	22 (37.9%)	24 (57.1%)	40 (40.4%)	65 (51.2%)	45 (49.5%)	
>70	144 (20.9%)	41 (15.0%)	13 (22.4%)	2 (4.76%)	26 (26.3%)	36 (28.3%)	26 (28.6%)	
Educational level (*N*=573)								<0.001***[,c]
None	32 (5.58%)	19 (7.60%)	0 (0.00%)	1 (2.38%)	12 (12.1%)	0 (0.00%)	NA	
Primary	164 (28.6%)	76 (30.4%)	13 (23.6%)	9 (21.4%)	33 (33.3%)	33 (26.0%)	NA	
Secondary	229 (40.0%)	78 (31.2%)	22 (40.0%)	19 (45.2%)	32 (32.3%)	78 (61.4%)	NA	
Tertiary	148 (25.8%)	77 (30.8%)	20 (36.4%)	13 (31.0%)	22 (22.2%)	16 (12.6%)	NA	
Post-menopause status (*N*=682)	450 (66.0%)	180 (66.4%)	32 (57.1%)	23 (59.0%)	59 (60.2%)	92 (72.4%)	64 (70.3%)	0.19
Archetype: (*N*=685)								<0.001***
Arch-1	102 (14.9%)	43 (15.8%)	8 (14.0%)	6 (14.3%)	9 (9.09%)	15 (11.8%)	21 (23.9%)	
Arch-2	84 (12.3%)	27 (9.93%)	3 (5.26%)	0 (0.00%)	17 (17.2%)	26 (20.5%)	11 (12.5%)	
Arch-3.1	68 (9.93%)	26 (9.56%)	3 (5.26%)	5 (11.9%)	13 (13.1%)	16 (12.6%)	5 (5.68%)	
Arch-3.2	198 (28.9%)	97 (35.7%)	18 (31.6%)	13 (31.0%)	17 (17.2%)	36 (28.3%)	17 (19.3%)	
Arch-4	54 (7.88%)	17 (6.25%)	4 (7.02%)	4 (9.52%)	9 (9.09%)	17 (13.4%)	3 (3.41%)	
Arch-5	58 (8.47%)	9 (3.31%)	11 (19.3%)	3 (7.14%)	16 (16.2%)	11 (8.66%)	8 (9.09%)	
Arch-6	51 (7.45%)	22 (8.09%)	6 (10.5%)	2 (4.76%)	9 (9.09%)	0 (0.00%)	12 (13.6%)	
Arch-7	70 (10.2%)	31 (11.4%)	4 (7.02%)	9 (21.4%)	9 (9.09%)	6 (4.72%)	11 (12.5%)	
Comorbidity[d] (*N*=666)	282 (42.3%)	108 (39.6%)	16 (28.6%)	14 (33.3%)	58 (58.6%)	56 (44.1%)	30 (43.5%)	0.00***
Tumour characteristics								
Ductal carcinoma in situ (*N*=689)	144 (20.9%)	43 (15.8%)	8 (13.8%)	12 (28.6%)	19 (19.2%)	15 (11.8%)	47 (52.2%)	<0.001***
Invasive ductal carcinoma (*N*=689)	495 (71.8%)	190 (69.6%)	40 (69.0%)	28 (66.7%)	84 (84.8%)	102 (80.3%)	51 (56.7%)	<0.001***
Invasive lobular carcinoma (*N*=688)	66 (9.59%)	19 (6.96%)	10 (17.2%)	6 (14.3%)	6 (6.06%)	10 (7.87%)	15 (16.9%)	0.01**
Other carcinoma (*N*=689)	32 (4.64%)	21 (7.69%)	0 (0.00%)	3 (7.14%)	0 (0.00%)	0 (0.00%)	8 (8.89%)	<0.001***
Positive oestrogen receptor status (*N*=656)	568 (86.6%)	238 (87.2%)	48 (82.8%)	38 (90.5%)	80 (83.3%)	99 (84.6%)	65 (92.9%)	0.42
Positive progesterone receptor status (*N*=655)	509 (77.7%)	221 (81.0%)	39 (68.4%)	32 (76.2%)	73 (76.0%)	82 (70.1%)	62 (88.6%)	0.02**
Positive HER2 receptor status: (*N*=674)	75 (11.1%)	23 (8.46%)	5 (8.62%)	7 (16.7%)	20 (20.8%)	16 (13.7%)	4 (4.49%)	<0.001***
Treatment characteristics								
Surgery: (*N*=688)								<0.001***
BCS	482 (70.1%)	214 (78.7%)	36 (63.2%)	29 (69.0%)	61 (61.6%)	81 (63.8%)	61 (67.0%)	
BCS with mammoplasty	33 (4.80%)	4 (1.47%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	29 (22.8%)	0 (0.00%)	
Mastectomy without immediate reconstruction	67 (9.74%)	11 (4.04%)	11 (19.3%)	3 (7.14%)	17 (17.2%)	15 (11.8%)	10 (11.0%)	
Mastectomy with immediate reconstruction	106 (15.4%)	43 (15.8%)	10 (17.5%)	10 (23.8%)	21 (21.2%)	2 (1.57%)	20 (22.0%)	
Surgery to axilla (*N*=686)								<0.001***
None	58 (8.45%)	27 (9.89%)	6 (10.5%)	4 (9.52%)	0 (0.00%)	6 (4.72%)	15 (17.0%)	
Sentinel lymph node biopsy	335 (48.8%)	198 (72.5%)	39 (68.4%)	24 (57.1%)	23 (23.2%)	0 (0.00%)	51 (58.0%)	
Axillary clearance	133 (19.4%)	48 (17.6%)	11 (19.3%)	13 (31.0%)	18 (18.2%)	22 (17.3%)	21 (23.9%)	
Axillary sampling	160 (23.3%)	0 (0.00%)	1 (1.75%)	1 (2.38%)	58 (58.6%)	99 (78.0%)	1 (1.14%)	
Radiotherapy (*N*=690)	567 (82.2%)	239 (87.5%)	42 (72.4%)	39 (92.9%)	66 (66.7%)	118 (92.9%)	63 (69.2%)	<0.001***
Chemotherapy (*N*=690)	248 (35.9%)	88 (32.2%)	22 (37.9%)	20 (47.6%)	43 (43.4%)	50 (39.4%)	25 (27.5%)	0.07
Hormonal therapy (*N*=690)	563 (81.6%)	231 (84.6%)	51 (87.9%)	37 (88.1%)	81 (81.8%)	97 (76.4%)	66 (72.5%)	0.04**
Targeted therapy (*N*=678)	66 (9.73%)	19 (7.01%)	5 (8.77%)	7 (20.6%)	18 (18.2%)	14 (11.0%)	3 (3.33%)	0.00***

Notes: BCS, breast cancer surgery; *N*, available sample size; NA, not applicable.

a: Percentages calculated excluding missing data.

b: *P*-value corresponding to the chi-square test.

c: Site F was excluded due to missing information.

d: Comorbidity has been defined as a binary variable equal to 1 if patient presents a comorbidity.

***: $P < 0.01$; **: $P <0.0$.

CRO assessed, *serious treatment-related complications* has shown significant differences, being *site* E the best performer and presenting a lower risk of complications (odds ratio = 0.26; 95% confidence interval = 0.07–0.73). Benchmarking of the health outcome indicators is shown in figure 1.

Half of the care-process indicators, *single operation for invasive cancer, single operation for DCIS*, and *appropriate chemotherapy*, have not shown statistically significant differences across healthcare centres. The other four care-process indicators, *immediate reconstruction, wound infection, appropriate time to surgery* and *sentinel node biopsy*, have shown a best/worst performer (e.g. *site E* showed a lower rate of *immediate reconstruction*).

All economic indicators have shown significant differences across sites. However, no best/worst performers have been identified since the economic performance should be interpreted along with the health outcomes performance.

Benchmarking of the care-process and economic indicators are shown in figure 2.

Discussion

VBHC proposes a strategic framework to manage healthcare systems connecting health care-process and economic outcomes to determine the value of the care. The VBHC implies systematically collecting and analysing PROs, CROs, CAPROs and EROs. To the best of our knowledge, this is the first VBHC coordinated experience across healthcare centres in different healthcare systems, collecting PCOs and providing evidence of health, care-process and economic dimensions to be used with a VBHC benchmarking intention. This paper aims to provide a proof-of-concept of the feasibility of an international VBHC in breast cancer to move toward a VBHC delivery system.

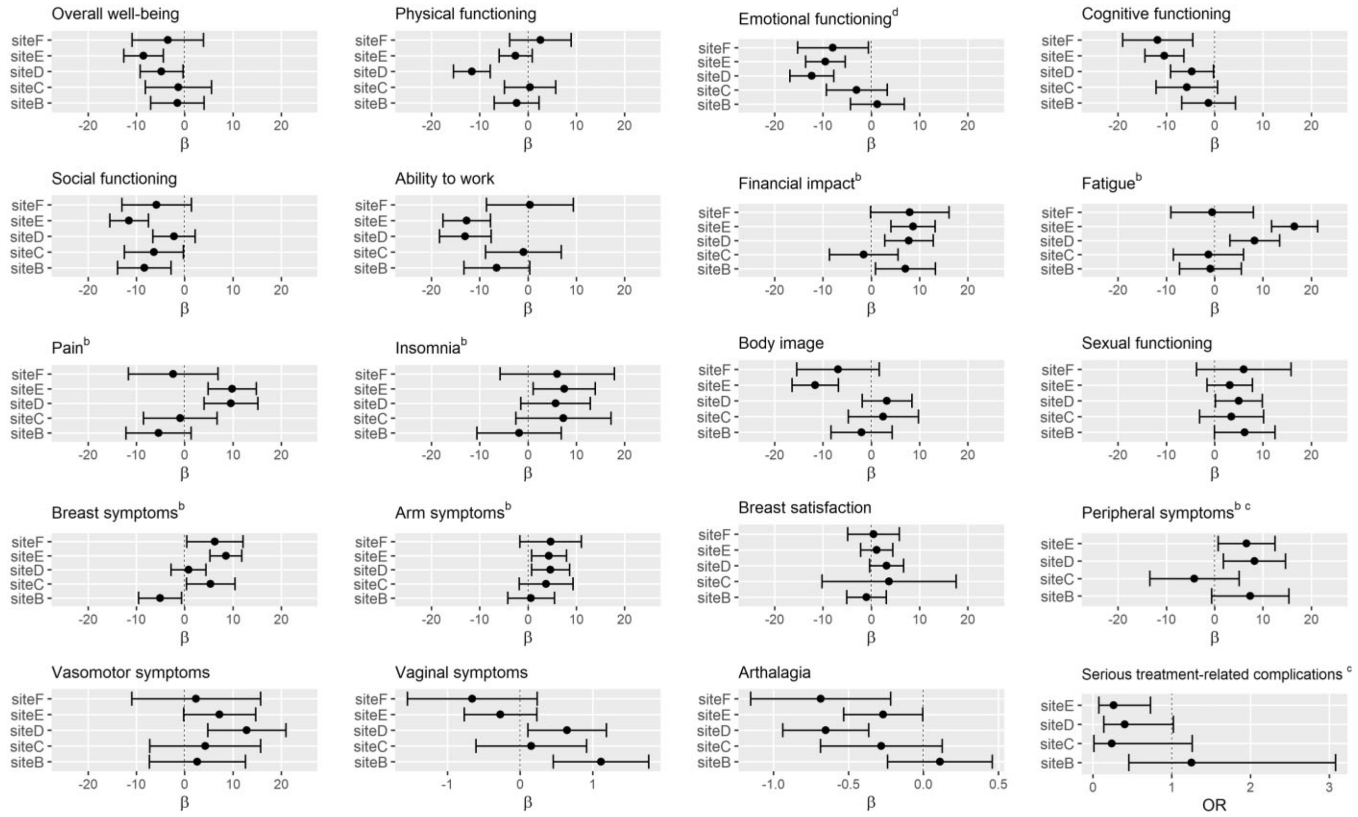

Figure 1 Benchmarking of health outcomes indicators.[a] IRR, incidence rate ratio; OR, odds ratio

[a]Site A is the reference site. Single operation rate for invasive cancer, single operation rate for DCIS, appropriate chemotherapy rate and appropriate time to surgery/treatment are care-process indicators (CAPROs); Specialist visits, Diagnostic tests, hospital length of stay, patient rate undergoing major outpatient surgery, day hospital sessions, patient treated with new therapies under chemotherapy are economic indicators (EROs). [b]No economic information available from site E. [c]Since no variability in outcome values was observed, Site E is omitted. [d]Since no variability in outcome values was observed, Site F is omitted. [e]Since no variability in outcome values was observed, Site C is omitted.

Following a two-step mixed-method approach based on a Delphi panel process[32] as qualitative methods, and clustering techniques and regression models as quantitative methods, this paper provides a benchmarking across six-European healthcare centres in three different countries in the context of the VOICE community.

Statistically significant differences across sites were found in most health outcomes, care-process indicators and economic indicators.

It is noted that classical health indicators such as *overall survival*, *disease-free-survival*, and *loco-regional recurrence*, and the care-process indicator *appropriate follow-up* were not assessed since the 6-month follow-up of the study made difficult their accurate interpretation. Besides, as the dependent variable of some regression models did not show variability for specific healthcare centres (e.g. healthcare centre C showing no wound infections), regression models were then run accordingly, removing the corresponding healthcare centre. A similar process occurred for the covariate of *archetype*. When an indicator did not apply for a specific archetype, the former was automatically removed from the regression models (e.g. *Archetype Arch3.2* with no chemotherapy was excluded from the regression model of the indicator *Appropriate chemotherapy rate*).

This study is not free of limitations. Patients with severe mental health problems or cognitive disorders were not excluded, this might alter the self-reported outcomes at some point; however, due to the low prevalence of these medical conditions, no significant impact was expected on the regression results. The ICHOM standard set proposes to collect PROs, CROs and CAPROs at several time periods. The VOICE community only collected PROs, CROs and CAPROs at 6 months, the first time period proposed by ICHOM. This short follow-up in the context of breast cancer might not be considering the entire patient pathway for some patients, which might bias the results. Since ICHOM periodically reviews their standard sets, the data dictionary version to collect the PROs and CROs used in this paper is not the latest. In terms of economic indicators, no best/worst performers have been identified since the economic performance should be interpreted along with the health and care-process outcomes performance. In other words, a pilot site showing the lowest *hospital length of stay* might be naively interpreted as the best performer, whether identified as the worst performer in specific health outcomes or care-process indicators. This joint interpretation of health, care-process and economic outcomes is currently a primary challenge in VBHC that the further research presented in this paper has planned to overcome. Furthermore, differences in the unit cost estimation methods of the healthcare resource should be explored in further research to allow for a reasonable benchmarking across sites. Benchmarking on resource use as economic indicators was considered informative enough for best practices sharing intention, and improving healthcare delivery processes. In this paper, healthcare costs were analyzed to be ultimately used only for the benefit of each healthcare centre. Besides, economic information was not reported by one healthcare centre. For certain indicators such as *serious treatment-related complications* and *single operation for DCIS*, confidence intervals are large for some particular sites because the eligibility patient's size is statistically small. Therefore, their interpretation should be taken with caution. The lack of collected PREs in this research is also considered a limitation.[19]

Efforts to bring the VBHC into routine clinical practice are a current challenge in healthcare centres. These efforts entail a paradigm shift across healthcare policies, information systems, healthcare professionals, and patients.

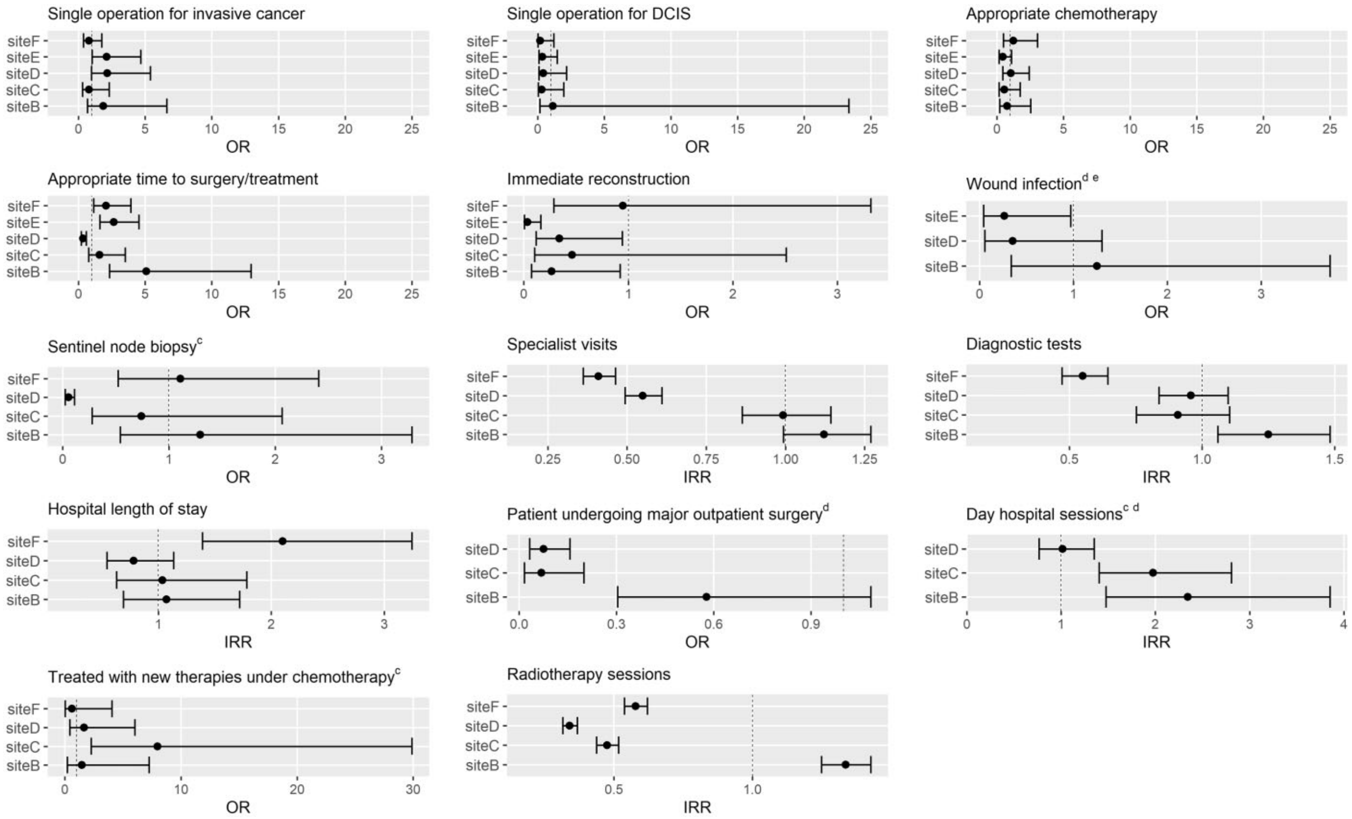

Figure 2 Benchmarking of care-process and economic indicators.[a,b] IRR, incidence rate ratio; OR, odds ratio

[a]Site A is the reference site. Single operation rate for invasive cancer, single operation rate for DCIS, appropriate chemotherapy rate and appropriate time to surgery/treatment are care-process indicators (CAPROs); Specialist visits, Diagnostic tests, hospital length of stay, patient rate undergoing major outpatient surgery, day hospital sessions, patient treated with new therapies under chemotherapy are economic indicators (EROs). [b]No economic information available from site E. [c]Since no variability in outcome values was observed, Site E is omitted. [d]Since no variability in outcome values was observed, Site F is omitted. [e]Since no variability in outcome values was observed, Site C is omitted.

Although the VOICE community has been conceived as a study pilot, the VBHC pathway is needed for a systematic collection of health, care-process and economic information to move towards a VBHC delivery system including the patient perspective, implementing the shared decision-making as routine in clinical practice, and reshaping the reimbursement systems used by the healthcare managers.

This paper provides empirical guidance to move forward with VBHC, overcoming the need for a standard set of EROs and their corresponding measure, and the absence of a validated set of indicators to benchmark. Both standard sets might be of use to other healthcare centres interested in improving their quality of breast cancer care based on the principles of VBHC. Furthermore, this paper provides a benchmarking across six-European healthcare centres on health outcomes, care-process and economic indicators based on robust methods sidestepping from the traditional descriptive analysis. Benchmarking should provide useful information to identify best practices and improve quality of healthcare providers. The use of observational data based on systematic data collection turns up as a crucial issue to move toward a VBHC delivery system. Furthermore, the methods of this research might interest researchers to move forward with VBHC in other medical conditions. The challenge lies in the development of methodological pathways to properly use the benchmarking information to share best clinical practices and improve the healthcare.

Supplementary data

Supplementary data are available at *EURPUB* online.

Acknowledgements

The authors thank Dr Nerea González for her worthwhile discussions on the DELPHI methodology. Members of the VOICE study group not listed among the co-authors are Iratxe Vázquez Lerma, Izaskun Artola Irazabal, Aitor Fernandez de Larrinoa Santamaría, Andere Frias Capanaga, Jon Cacicedo, Eduardo Vicario Elorduy, Teresa Acaiturri-Ayesta, Elisa Gómez-Inhiesto, Valentina Danesi, Nicola Gentili, Andrea Roncadori, Fátima Hermoso Alarz, Karmele Imaz Iraola, Valerie Adam, Helene De Rijck, Ellen Everaert, Esmeralda Barreira, Pedro Medeiros and Emanuel Barros.

Funding

The VOICE Project, or 'Value-Based Healthcare for Outcomes in Breast and Lung Cancer in Europe', coordinated by the Kronikgune Institute, has the collaboration of Roche Farma, SA and Roche Diagnostics SLU.

Disclaimer

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of interest: None declared.

Data availability

The data underlying this article cannot be shared publicly due to privacy of individuals that participated in the study.

Key points

- There is a lack of VBHC benchmarking in breast cancer, which has become a challenge following a threefold reason: (i) the low adherence to systematically measuring outcomes by healthcare centres, (ii) the lack of a standard set of Economic-Related Outcomes (EROs) and (iii) the absence of a validated set of breast cancer indicators to benchmark.
- This paper provides empirical guidance to move forward with VBHC, overcoming the need for a standard set of EROs, and the absence of a validated set of indicators to benchmark.
- The paper delivers benchmarking across six-European healthcare centres on health, care-process and economic indicators based on robust methods sidestepping from traditional descriptive analysis.
- This paper aims to provide a proof-of-concept of the feasibility of an international VBHC benchmarking of health, care-process and economic outcomes in breast cancer, with the ultimately aim of being used to share best practices with a data-driven approach among healthcare organizations from different health systems.
- Methods of this research might be of interest to move forward with VBHC in other medical conditions.

References

1 Grilli R, Espin J, Florindi F, Lorenzo FD. *Policy Paper on Enhancing the Value of Cancer Care Through a More Appropriate Use of Healthcare Interventions*. Cancer Control Joint Action 2017. https://cancercontrol.eu/archived/uploads/PolicyPapers27032017/Policy_Paper_3_Enhancing.pdf (10 April 2023, date last accessed).

2 Porter ME. What is value in health care? *N Engl J Med* 2010;363:2477–81. doi: 10.1056/NEJMp1011024

3 EIT Health. *Implementing Value-Based Health Care in Europe: Handbook for Pioneers* (Director: Gregory Katz). 2020.

4 Cossio-Gil Y, Omara M, Watson C, et al. The roadmap for implementing value-based healthcare in European University Hospitals—consensus report and recommendations. *Value Health* 2022;25:1148–56.

5 Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. *Acad Med* 2020;95:682–5.

6 Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: global statistics for 2020 and 2040. *Breast* 2022;66:15–23.

7 Mokhtari-Hessari P, Montazeri A. Health-related quality of life in breast cancer patients: review of reviews from 2008 to 2018. *Health Qual Life Outcomes* 2020;18:338.

8 Luengo-Fernandez R, Leal J, Gray A, Sullivan R. Economic burden of cancer across the European Union: a population-based cost analysis. *Lancet Oncol* 2013;14:1165–74.

9 Ong WL, Schouwenburg MG, van Bommel ACM, et al. A standard set of value-based patient-centered outcomes for breast cancer: the International Consortium for Health Outcomes Measurement (ICHOM) Initiative. *JAMA Oncol* 2017;3:677–85.

10 Biganzoli L, Marotti L, Hart CD, et al. Quality indicators in breast cancer care: an update from the EUSOMA working group. *Eur J Cancer* 2017;86:59–81.

11 Atsma F, Elwyn G, Westert G. Understanding unwarranted variation in clinical practice: a focus on network effects, reflective medicine and learning health systems. *International Journal of Quality in Health Care* 2020;32:271–4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270826/pdf/mzaa023.pdf (20 April 2023, date last accessed).

12 ICHOM Breast Cancer Group. *Non-Metastatic Breast Cancer*. 2017. www.ichom.org/medicalconditions/breast-cancer (19 May 2019, date last accessed).

13 Trebble TM, Hansi N, Hydes T, et al. Process mapping the patient journey: an introduction. *BMJ* 2010;341:c4078.

14 Panella M, Marchisio S, Di Stanislao F. Reducing clinical variations with clinical pathways: do pathways work? *Int J Qual Health Care* 2003;15:509–21.

15 Drummond MF, Sculpher MJ, Claxton K, et al. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press, 2015.

16 Chapko MK, Liu CF, Perkins M, et al. Equivalence of two healthcare costing methods: bottom-up and top-down. *Health Econ* 2009;18:1188–201.

17 Jones J, Hunter D. Qualitative research: consensus methods for medical and health services research. *BMJ* 1995;311:376–80.

18 Williams KE, Sansoni J, Morris D, Thompson C. A Delphi study to develop indicators of cancer patient experience for quality improvement. *Support Care Cancer* 2018;26:129–38. doi:10.1007/s00520-017-3823-4.

19 Minvielle E, Fierobe A, Fourcade A, et al. The use of patient-reported outcome and experience measures for health policy purposes: a scoping review in oncology. *Health Policy* 2023;129:104702.

20 Likert R. A technique for measurement attitudes. *Arch Psuchol* 1932;55–65.

21 Husson F, Josse J, Pages J. Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? 2010;17.

22 Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Soft* 2008;25: doi:10.18637/jss.v025.i01

23 Iezzoni LI. The risks of risk adjustment. *JAMA* 1997;278:1600–7.

24 Iezzoni LI, Ash AS, Shwartz M, et al. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health* 1996;86:1379–87.

25 Dubois RW, Brook RH, Rogers WH. Adjusted hospital death rates: a potential screen for quality of medical care. *Am J Public Health* 1987;77:1162–6.

26 Lovaglio PG. Benchmarking strategies for measuring the quality of healthcare: problems and prospects. *ScientificWorldJournal* 2012;2012:606154.

27 El-Saed A, Balkhy HH, Weber DJ. Benchmarking local healthcare-associated infections: Available benchmarks and interpretation challenges. *J Infect Public Health* 2013;6:323–30.

28 Vincent BM, Wiitala WL, Luginbill KA, et al. Template matching for benchmarking hospital performance in the veterans affairs healthcare system. *Medicine (Baltimore)* 2019;98:e15644.

29 Ang D, McKenney M, Norwood S, et al. Benchmarking statewide trauma mortality using Agency for Healthcare Research and Quality's patient safety indicators. *J Surg Res* 2015;198:34–40.

30 Jones A. *Applied Econometrics for Health Economists: A Practical Guide*. London: CRC Press; 2007.

31 Cameron AC, Trivedi PK. Regression-based tests for overdispersion in the Poisson model. *J Econom* 1990;46:347–64.

32 De Meyrick J. The Delphi method and health research. *Health Educ* 2003;103:7–16.