Abstract

Herceptin (Trastuzumab) is a widely used and effective drug for the treatment of Her2+ breast cancer but its cardiotoxic side effects require regular monitoring by echocardiography. A 10% reduction in left ventricular ejection fraction can lead to suspension of treatment and therefore has significant implications for patient prognosis in terms of cardiac and cancer outcomes. Assessment of LV function by conventional 2D biplane method of discs (2DEF) has limitations in accuracy and reproducibility. Global longitudinal strain (GLS) is becoming more widely available and user friendly. It has been shown to demonstrate myocardial damage earlier in treatment than 2DEF, allowing the option of pharmacological intervention at a pre-clinical stage and preventing the interruption of Herceptin. This study compares the reproducibility of GLS with that of 2DEF in a routine clinical environment. Fifty echocardiograms performed on female patients undergoing Herceptin treatment were used to measure both 2DEF and GLS within the recommended standard appointment time of 40 min. The data were re-measured (blind) by the same operator a minimum of 14 days later to determine intra-operator variation. These data were also measured by a second operator (blind), to assess inter-operator variation. Analysis by direct comparison, intra-class correlation (ICC), coefficient of variation (CV) and Bland–Altman plots demonstrated that GLS is a more reproducible measurement than 2DEF. This is important to prevent clinical decisions being erroneously based on variation in operator measurement. The investigation also shows that with advances in machine software this is a practical addition to routine assessment rather than merely a research tool.

Introduction

Breast cancer is the most common cancer in the UK with over 53,000 cases per year accounting for 15% of all cancers diagnosed (http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer, accessed April 2016). Of these 20–30% involve the overproduction of protein human epidermal growth factor receptor-2 (HER-2), leading to rapid cell growth and tumour formation (1) and associated with adverse prognosis and shorter survival times (2, 3). Herceptin is a monoclonal antibody treatment used in Her2+ve breast cancer either as monotherapy or more commonly, in combination with anthracycline. The cardiotoxic side effects of Herceptin are well established and therefore cardiac monitoring, usually by echocardiography, is mandatory.

The assessment of ejection fraction by Simpson’s biplane which measures changes in volume has limitations in sensitivity and reproducibility as a result of geometric modelling, inadequate visualisation of the left ventricular apex and inherent measurement variability. This often hinders the detection of small changes in contractility (4, 5, 6, 7, 8). The 95% confidence intervals of ejection fraction by Simpson’s biplane vary between 8.9% (9) and 10.8% (5) which is within the limits of a
10% reduction in left ventricular systolic function and so could trigger potentially inappropriate treatment suspension of Herceptin if overall ejection fraction dropped to less than 55%.

Speckle tracking echocardiography (STE) assesses the longitudinal, radial and circumferential regional deformation or ‘myocardial strain’. Global longitudinal strain (GLS) uses apical images which are more easily and reproducibly obtained than the parasternal images used for radial or circumferential assessment (10). Compared with 2D ejection fraction GLS has been shown to identify subclinical cardiac dysfunction at an earlier stage in a number of studies with Herceptin treatment (10, 11, 12).

Initially the measurement of GLS was a time-consuming and complex process requiring off-line analysis and confined to research laboratories and specialist centres. However, advances in technology and software have resulted in the development of more automated systems with anatomical recognition now available on most mid and higher specification echo machines and some compact systems. This reduces the need for operator input thereby decreasing measurement variability.

Analysis can be made ‘on-cart’ at the time of the procedure. Operator input is restricted to ensuring that the automated software has correctly identified anatomical markers and tracked the cardiac cycle effectively. Minor adjustments can be made manually if the software has misinterpreted artefact as an anatomical structure or if an internal structure such as a papillary muscle has obscured the endocardial border. A more automated system should lead to a reduction in intra- and inter-operator variability and should be quick and practical enough to apply in a busy clinic without impacting on appointment times. The 2014 ASE/EACI Expert Consensus Statement for evaluation of patients during cancer therapy states that the lack of published data regarding the reproducibility of GLS at non-academic centres or community hospitals is one of the limitations of GLS measurement (8).

Methods

Cohort selection

Initial selection All patients attending the Herceptin clinic between 01/02/2015 and 31/07/2015 for assessment of cardiac function prior to or during Herceptin treatment were considered potential candidates for this project where operators and machines were available. This gave a maximum cohort of 75 patients over a 6-month period. James Cook University Hospital Research and Development Audit committee concluded this work fell into the category of Service Evaluation and posed no unacceptable governance or ethical issues and as such formal ethical approval and specific patient consent for participation was not required.

Exclusions Patients with an irregular heart rhythm, conduction abnormalities or pacemakers were excluded because in-coordinate contraction or inconsistent cardiac cycle length can give rise to variations in ejection fraction or peak systolic strain times (n=9). Those with suboptimal image quality resulting in poor endocardial definition precluding ejection fraction by Simpson’s Biplane method of discs were also excluded from this study as the objective was direct comparison of two techniques. No patients presented with significant valve disease.

Final cohort The final cohort was 43 females with mean age 55 (range 24–56) years at various stages of Herceptin treatment. Seven patients were scanned twice over the 6-month period giving a total number of 50 scans for analysis.

Equipment used

All patients were scanned using a Philips ‘Epiq 7’ cardiac ultrasound machine with Automated Cardiac Motion Quantification (aCMQ) software installed. The local optimised pre-set protocol was used and further optimisation of images with each individual patient was carried out in line with British Echocardiography Society Guidelines to achieve the best visualisation of the myocardium.

Study aims

The aim of this study was to compare the reproducibility of left ventricular function assessment by 2DEF with GLS in a real-time outpatient clinic in the setting of Herceptin therapy.

Staff involved

Operator 1 AK: BSE accredited cardiac physiologist with >10-year experience in echocardiography.
Operator 2 EL: BSE accredited cardiac physiologist with >10-year experience in echocardiography.

Supervising Cardiac Consultant MJS: Imaging specialist with BSE accreditation.

Data acquisition

A full standard dataset of images was obtained for each patient in line with 2011 BSE minimum standard dataset to allow a full and comprehensive assessment of cardiac chamber sizes, wall thickness, anatomical features, heart valve function, right and left ventricular systolic and diastolic function, and the presence of pericardial fluid. The images required specifically for analysis of left ventricular function (apical 4 chamber, 2 chamber and 3 chamber views) were recorded for 2 cardiac cycles in held respiration to minimise translational errors and avoid possible aberrant machine clipping of a single cardiac cycle. The entire dataset was stored in the Philips echo machine hard drive and an identical copy was also transferred to the current hospital database for routine analysis and reporting back to the requesting physician using the in-house reporting package ‘Prosolv’ (Problem Solving Concepts Cardiovascular Analyser 3.5, FUJIFILM Medical USA, Inc., Stamford, CT, USA). All image optimisation and dataset acquisition were performed by Operator 1 (AK).

Data analysis

Measurement of 2DEF Calculation of ejection fraction was made using the stored image loops of the apical 4 chamber and apical 2 chamber views. Measurements were performed on the in-house hospital reporting system ‘Prosolv’ using the recommended Simpson’s biplane method of discs in line with current routine practice for calculation of ejection fraction. Measurements were then deleted from the machine in order that re-measurement was blinded to the original analysis. These initial 2DEF and GLS measurements were made and recorded during the standard allocated appointment time for the patient (40 min).

Assessment of intra-operator variability The study date, patient ID number and image identification numbers used in the first analysis were used by the original operator (AK) to reload the patient dataset on the Philips Epiq 7 machine and re-measure the GLS. This was done a minimum of 14 days after the first analysis to avoid any bias from remembering the previous result. The original operator (AK) also reloaded the same patient dataset on ‘Prosolv’ and re-measured ejection fraction using the same method (Simpson’s biplane method of discs). Measurements were then deleted to blind the study as before.

Assessment of inter-operator variability A second operator (EL) used the study date, patient ID number and image identification numbers from the first analysis to reload the patient dataset and measure the GLS on the Philips Epiq 7 machine. Any adjustments to GLS automated software judged necessary by the second operator were made and the GLS value recorded on the third data worksheet. The second operator then reloaded the same patient dataset on ‘Prosolv’ and measured ejection fraction using Simpson’s biplane method of discs on the same images as the original analysis. Frame selection and manual tracing of the endocardial border were made at the second operator’s discretion.

Statistical analysis

Statistical analysis was done using Medcalc software to calculate intra-class correlation (ICC), coefficient of variation (CV) and Bland–Altman plots. ICC is a measure of consistency or reliability of quantitative measurements made by different observers measuring the same quantity. Results are given as a score between 0.0 and 1.0, where the extent of variation between rater measurements is inversely proportional to the score. Thus a value of 1.0 means there is zero measurement error. Deciding what constitutes clinically acceptable reliability is subjective but generally a higher value is preferred. CV is a measure of variability in relation to the mean given as a ratio of standard deviation to the mean allowing two ratios to be compared. It is useful in comparing one data series to another. The variable with
the lower coefficient of variation is the less dispersed and thus more reproducible. Bland–Altman plots are used to evaluate the agreement among two different measurement techniques and to assess the repeatability of a method. The closer the means are to zero, the greater the reproducibility.

Variability was also grouped in terms of significance of outcome for 2DEF and GLS measurement to demonstrate the number of patients for whom the difference in operator measurement would have triggered a change in clinical management.

Results

Results are summarised in Tables 1 and 2. Table 1 shows intra-operator reproducibility was good, with no clinically significant difference between first and second measurements of either 2DEF (means of 66.1 and 65.9 respectively) or GLS (means of −19.6 and −19.5). There did appear to be a 6% difference between operators (means 66 and 60) in 2DEF measurements but in contrast only a 0.2 absolute difference in GLS (means −19.6 and −19.4). Table 2 shows both intra-operator variability and inter-operator reliability were better with GLS than 2DEF (ICC 0.95 vs 0.76 and ICC 0.92 vs 0.66 respectively). However, the difference between intra-observer and inter-observer variability was less marked in measurement of GLS. Similarly, reproducibility was also better with GLS both within (CV 3.3% vs 6.3%) and between (3.97% vs 10.27%) observers. Intra-operator reproducibility was better than inter-operator reproducibility but the difference was smaller for GLS than for 2DEF.

Findings are supported by Bland–Altman plots in Figs 1 and 2 with lower mean values for GLS (−0.2 inter and −0.1 intra) than EF measurements (6.3 inter and 0.5 intra).

Impact on clinical management

Change in EF is generally expressed in absolute values i.e. a 10% change in EF would be from 60 to 50%. This study showed that 12 patients out of 50 showed a >10% change in EF when measured by 2 different operators. Of these, 6 would potentially have had their treatment suspended as this represented a drop to below an EF of 55%. Three patients had a >10% change in EF when re-measured by the same operator. One of these represented a drop to >55% and potential treatment suspension (Table 3).

Change in GLS is generally expressed in actual % change, where a 15% drop in GLS represents significance, no patients out of 50 showed a % change in excess of this and therefore none would have potentially had their treatment suspended (Table 4).

In our clinical practice all results which might result in a change in patient management are referred to a senior imaging cardiologist for review and side-by-side eyeball comparison to previous images in line with ASE and EACI guidelines. A final decision regarding treatment continuation, alteration or initiation of cardioprotective agent is made in consultation with the referring oncologist. The measurements in this study

### Table 1 Mean, 95% confidence intervals and standard deviation for each group.

<table>
<thead>
<tr>
<th>Groups compared</th>
<th>EF intra-operator variability</th>
<th>EF inter-operator variability</th>
<th>GLS intra-operator variability</th>
<th>GLS inter-operator variability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EFOP1a and EFOP1b</td>
<td>EFOP1a and EFOP2</td>
<td>GLSOP1a and GLSOP1b</td>
<td>GLSOP1a and GLSOP2</td>
</tr>
<tr>
<td>ICC</td>
<td>0.76</td>
<td>0.66</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>CV (%)</td>
<td>6.30</td>
<td>10.27</td>
<td>3.30</td>
<td>3.97</td>
</tr>
</tbody>
</table>

EFOP1a, Ejection fraction by Operator 1 first analysis; GLSOP1a, GLS by Operator 1 first analysis; EFOP1b, Ejection fraction by Operator 1 second analysis; GLSOP1b, GLS by Operator 1 second analysis; EFOP2, Ejection fraction by Operator 2; GLSOP2, GLS by Operator 2.
Figure 1
Bland–Altman plots showing inter-operator and intra-operator measurement of 2DEF.

Figure 2
Bland–Altman plots showing inter-operator and intra-operator measurement of GLS.

Discussion
This study demonstrated that GLS is a more reliable and reproducible method for assessing left ventricular function than 2D ejection fraction. This reduces the possibility of erroneous clinical decisions, potentially denying prognostic therapy based on operator variability rather than true clinical changes. Appropriate use of GLS would have indicated the discontinuation of Herceptin in none of the patients in this study compared with six patients based on 2DEF. Whilst it is preferable for serial measurements to be made by one operator this is not always practical in large departments. This study showed that the difference between intra- and inter-observer variability in the assessment of GLS is less marked compared with 2DEF, suggesting that there can be greater flexibility in the performance of such scans. Furthermore these studies were performed within the existing standard appointment slot, indicating that GLS can be easily and rapidly measured without impacting on the workflow or the need for more specialised offline equipment.

The benefits of Herceptin as a treatment for Her2+ve breast cancer are widely accepted and in 2010 it was also approved for Her2+ve metastatic adenocarcinoma of the stomach and gastro-oesophageal junction (2).
Given the significant reduction in mortality, recurrence and metastases shown by Herceptin treatment, accurate and reliable assessment of cardiac function is important to ensure that Herceptin is not inappropriately discontinued and that appropriate cardiac therapy to improve cardiac function is commenced in a timely fashion.

Comparison to previous studies

This study showed similar or better ICC and CV values compared with previous work measuring GLS using speckle tracking echocardiography where intra- and inter-operator variance of GLS were included. There was also better reproducibility with a single operator (i.e. less intra-operator variance) than with two operators (inter-operator variance) in accordance with our present study (Table 5).

Data analysis using the previous generation software was more time consuming as it was performed off-line and required greater operator input to identify anatomical markers such as mitral valve annulus and LV apex. Most were investigations into the diagnostic and prognostic value of GLS measurement in various cardiac pathologies and included reproducibility in a discrete number as part of the study.

Few studies have made a direct comparison to 2DEF alone, particularly in a real-life practical setting outside the research environment. Belghiti and coworkers (13) compared GLS, 2DEF and cineventriculography using older versions of automated GLS software with a different system (GE, workstation Echopac) and found GLS easy to apply and more reproducible than 2DEF but used a cohort of mixed gender, age and varying cardiac pathologies, including valve disease, which may have haemodynamic implications in EF assessment. Costa and coworkers (14) also examined mixed pathology patients and concluded GLS is highly reproducible in a ‘real-world’ setting but their study involved scanning all patients twice at the same appointment to include an assessment of vendor–vendor variability doubling their attendance time and did not compare GLS to 2DEF.

This study compares GLS measurement to current standard assessment of LV with 2DEF in patients undergoing serial echocardiography where identifying a trend of change is the main objective and therefore minimising operator variability is paramount. It does this in the setting of a routine clinic to test reproducibility of measurements under realistic circumstances and assess their value in the diagnostic evaluation of LV function.

Limitations

It is acknowledged that as only the measurements and not the image acquisition were repeated by Operator 1 and Operator 2 this was not a wholly true reflection of intra- and inter-operator variation. However, the primary aim of the study was to directly compare the method and application of software under current use and applying it to identical images operator technique is the only variable present. The measurement of GLS on-cart whilst 2DEF was determined off-line is another potential confounder but this adhered to the current departmental practice of EF measurement with Prosolv software in order to avoid the introduction of a second variable by using a less familiar on-cart software. Similarly, experienced operators reduced the likelihood of 2DEF measurement variability caused by inexperience.

This study did not use a defined classification of image quality but the same image was used for assessment by both methods as long as the quality was sufficient to allow 2DEF Simpson’s biplane as this would also allow satisfactory GLS analysis. The aCMQ software has built-in quality levels and will not display any section

Table 3  Variability in ejection fraction of 50 scans performed.

<table>
<thead>
<tr>
<th>EF variability</th>
<th>0–5%</th>
<th>6–10%</th>
<th>11–15%</th>
<th>&gt;15%</th>
<th>Total number of scans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-observer</td>
<td>36</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Inter-observer</td>
<td>22</td>
<td>16</td>
<td>7</td>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>

Values given are in absolute % change (i.e. a 10% change in EF would be from 60 to 50%).

Table 4  Variability in global longitudinal strain of 50 scans performed.

<table>
<thead>
<tr>
<th>GLS variability</th>
<th>0.0–5.0%</th>
<th>5.1–10.0%</th>
<th>10.1–15%</th>
<th>&gt;15.0</th>
<th>Total number of scans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-observer</td>
<td>39</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Inter-observer</td>
<td>31</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>50</td>
</tr>
</tbody>
</table>

Values given in actual % change (i.e. a 10% change in GLS would be −20.0% to −18.0%).
of the image that is not tracked to its pre-set standard. There were no instances where any segments were excluded by the software. This was used as a reflection of adequate image quality despite its lack of quantification. Alternative approaches to improving accuracy include use of IV contrast or 3D echo but both will be more time consuming for patient and operator. In addition, contrast involves cannulation, and 3D echo is more dependent on image quality than GLS.

This study used Philips Epiq 7 equipment and GLS calculation software and hence the results may not be transferable to other vendor’s equipment. There are a number of studies as recent as 2015 (15) that indicate small but significant differences in GLS measurement between vendors.

The lack of an established and accepted set of normal values for GLS is currently one of its biggest drawbacks. A number of studies have tried to establish normal values for GLS over a range of ages but the findings vary and the factors affecting GLS are contradictory. The HUNT study by Dalen and coworkers (16) in 2009 looked at 1266 healthy subjects and gave a normal value −17.4 (SD 2.3) for females and −15.9 (SD 2.3) for males and showed a decrease with age. Marwick and coworkers (17) gave the normal range as −18.6 ± 2.0 but concluded that weight, blood pressure and heart rate were influencing factors. The JUSTICE study of 2012 (18) looked at 817 healthy subjects and gave a figure between −16.4 and −23.4 as normal range but demonstrated variation between vendors only. A meta-analysis in 2013 (19) looked at a number of studies using both healthy subjects and those with a variety of cardiac pathologies, totalling 2597 subjects in all. This concluded a normal range of −15.9 to −22.1 with no age, gender or vendor bias but an influence by blood pressure. The latest study by Kocabay in 2014 (20) looked at 247 healthy subjects and concluded a lower limit of −18.5 for females, a higher value for males, no age or blood pressure bias but a possible influence of body mass index.

For these reasons, the American Society of Echocardiography and European Association were not prepared to give a lower cut-off value for GLS in their 2015 Guideline and Standards and stated only that a GLS of −20 should be considered normal. However, as this study is focused on relative change rather than deviation from an accepted norm, this lack of established normal values does not detract from the findings. In addition, it is acknowledged that GLS is not suitable for patients with irregular heart rhythms such as atrial fibrillation, frequent ectopy or pacing.

**Future considerations**

The introduction of new HER2+ve specific treatments such as Lapatinib and Kadcyla (3) mean that the numbers of patients requiring regular cardiac monitoring for Cancer Therapeutic Related Cardiac Dysfunction (CTRCD) is likely to increase. Anthracyclines such as Doxorubicin can also cause cardiac dysfunction, which may be permanent and irreversible. There is growing concern about late presentation CTRCD emerging months or years after treatment (21), which is a particular concern with paediatric cancers where treatment as a child or adolescent may have significant cardiac implications for much of the adult life span of patients. Data are emerging examining the prevention of cardiotoxicity using drugs such as angiotensin converting enzyme inhibitors (ACEIs), beta-blockers and angiotensin receptor blockers (ARBs). A meta-analysis by Kalam and Marwick in 2013 (22) demonstrated considerable success in the prophylactic use of ACEIs, ARBs and beta-blockers for the prevention of heart failure in anthracycline treatment. Early recognition of sub-clinical dysfunction by GLS measurement may...
enable prompt consideration of appropriate therapy to potentially ameliorate toxic effects of chemotherapy. A reduction in GLS by 15% from baseline, particularly one accompanied by a rise in Troponin I is a strong indicator of subclinical dysfunction and should initiate a discussion between cardiology and oncology regarding the risk and benefits of continuing chemotherapy and the initiation of treatment for heart failure. However at present there is no clear evidence that chemotherapy should be withheld. ASE/EACI recommendations include echocardiographic assessment after each further treatment cycle to identify dysfunction and to inform the patient of the risks of continuation (8).

Tracking of LV function using GLS may also be useful in the monitoring of severe but asymptomatic valve disease and in the earlier diagnosis of inherited cardiac disorders such as hypertrophic cardiomyopathy (HCM) and arrhythmogenic right ventricular cardiomyopathy (ARVC). Studies by Saito and coworkers (23) and Ozawa and coworkers (24) both showed reduced GLS measurements in known HCM patients with normal 2D EF.

The demonstration of reduced LVEF without a concomitant change in GLS in a group of 6 patients is worthy of further study as chemotherapy was not discontinued. Prospective follow-up of cardiac outcomes in a larger cohort of such patients would lend weight to the importance of the assessment by GLS.

### Conclusion

This study demonstrates that GLS is a more reproducible method for assessing left ventricular systolic function than 2D ejection fraction, adding security to the clinical decision-making process regarding continuing Herceptin therapy, an agent which improves prognosis in breast cancer. The difference between intra- and inter-observer variability in the assessment of GLS is less marked compared with 2D EF suggesting that such scans could be performed by more than one operator without reducing accuracy and reproducibility, allowing greater flexibility in busy departments. All assessments were performed within the existing standard appointment slot indicating that it is an accessible tool in non-specialist centres and community settings as well as in tertiary units.

### Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.


Received in final form 26 August 2016
Accepted 14 September 2016
Accepted Preprint published online 14 September 2016