Background
Since 2004, National Institute for Health and Care Excellence (NICE) methodological guidance has contained preferred methods for measuring patient benefits as utility scores in economic evaluations conducted for NICE. The preferred methods are as follows:

• EQ-5D is the preferred utility instrument
• Quality of life questionnaires should be completed by patients or carers if patients completions is unfeasible
• The general public should value health states
• Where EQ-5D is not directly available, mapping algorithms from other measures should be considered (added in 2008 guidance)
• Where EQ-5D and mapping are not available, validated methods using the time trade-off method of quality of life valuation are preferred with full health as the upper anchor (since 2008)
• Justifications should be provided where methods deviate from the NICE Reference Case

This study evaluates whether methodological guidance has produced utility score consistency among NICE single technology appraisals (STAs) using STAs of breast cancer as a case study.

Objective
The aim of the research was to assess whether utility scores used in breast cancer STAs were consistent with the NICE Reference Case, with original sources and between STAs.

Methods
A review of all published breast cancer STAs was undertaken using all publicly available STA documents. STAs were divided into early breast cancer (EBC) and metastatic breast cancer (MBC) publications. Utility scores were assessed for consistency with NICE preferred methods, with sources, and across STAs. Academic assessment group work was examined to show whether they emphasized NICE preferred methods.

Results

EBC – NICE Reference Case consistency:
• All three EBC STAs compared against the 2004 NICE Reference Case.
• In the three STAs there were 27 descriptions of health states
  • 8 redacted as academic or commercial in confidence (30%)
  • 2 used EQ-5D but the scores were redacted (7%)
  • 7 had patient filling questionnaires (26%)
  • 8 utility states were valued by the general public (30%)
  • 10 used standard gamble or time trade-off methods (37%)
• Multiple methods of utility measurement were commonly used in the same study, without adjustment for differences in method

MBC – NICE Reference Case consistency:
• Six MBC STAs, 2004 Reference case only relevant for TA 116
• 107 descriptions of health states and decrements, only 4 source papers
• 3 of 4 source papers used regression, 1 regression included EQ-5D
• 1 source paper included patients filling questionnaires
• 3 had valuation by the general public
• 0 source papers used mapping
• 5 of 6 MBC STAs had methodological guidance advising mapping, why no STAs used mapping algorithms is unclear
• All 6 STAs used quality of life instruments that were mappable to EQ-5D
• Only one Evidence Review Group requested clarification from the manufacturer about the use of mapping algorithms

Consistency of utility values across STAs:
• Utility values for same health state or adverse event often inconsistent, especially between EBC and MBC; many different assumptions

Table 1: Consistency of Early Breast Cancer Utility Sources with the NICE Reference Case (2004)

<table>
<thead>
<tr>
<th>Utility Source</th>
<th>Source papers</th>
<th>Utilities measured</th>
<th>Valuation by the general public</th>
<th>EQ-5D used</th>
<th>Preference based method (not rating scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hébrard and Smith (1993)</td>
<td>107, 109</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sorensen et al. (2004)</td>
<td>108</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Drea et al. (2004)</td>
<td>109</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Brown and Hutton (1998)</td>
<td>109</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Austin et al. (1996)</td>
<td>109</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2: Consistency of Metastatic Breast Cancer Utility Sources with the NICE Reference Case

<table>
<thead>
<tr>
<th>Utility Source</th>
<th>Source papers</th>
<th>Utilities measured</th>
<th>Valuation by the general public</th>
<th>EQ-5D used</th>
<th>Preference based method (not rating scale); TTO post 2008</th>
<th>Mapping use?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lloyd et al. (2006)</td>
<td>2004</td>
<td>116</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Lloyd et al. (2008)</td>
<td>2008</td>
<td>214, 239, 250, 263, 295</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cooper et al. (2003)</td>
<td>2008</td>
<td>214</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hutton et al. (1996)</td>
<td>2008</td>
<td>250</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Peacock et al. (2000)</td>
<td>2008</td>
<td>214</td>
<td>Partially</td>
<td>Partially</td>
<td>Partially</td>
<td>Partially</td>
</tr>
</tbody>
</table>

Conclusion
Breast cancer STAs have shown a broad lack of compliance with preferred methods of measuring utility scores in the NICE Reference Case and a lack of explanation regarding this lack of compliance. Further, variability in utility scores due to differing methods and/or assumptions in the various STAs, even when these STAs have derived their utility scores from the same source highlights the potential for inconsistency in NICE recommendations. Evidence Review Groups performed poorly on critiquing and highlighting this inconsistency. The flowchart above can help encourage manufacturers to use appropriate utility measurement methods, and provide Evidence Review Groups with an easy to use assessment tool to assess consistency.