Economic evaluation in mental healthcare
Stant, Dennis.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 6

Cost-effectiveness analysis in severe mental illness: outcome measures selection

Abstract

Background: Most economic evaluations conducted in mental healthcare did not include widely recommended preference-based health outcomes like the QALY (Quality-Adjusted Life Years). Instead, studies were mainly designed as cost-effectiveness analyses that included single outcome measures aimed at a specific aspect of health.

Aims of the study: To raise awareness about the potential problems related to the selection of outcome measures for economic studies in patient populations with severe mental illness. Furthermore, to make suggestions that may prevent these problems in future economic evaluations.

Methods: Data from a previously conducted economic evaluation assessing the cost-effectiveness of the HIT (Hallucination focused Integrative Treatment) intervention in patients with severe mental illness was used for the analyses presented in the current paper. Economic analyses based on the results of the selected primary health outcome (Positive And Negative Syndrome Scale: PANSS) were compared with results based on various other health outcomes assessed during the study, including QALYs.

Results: No relevant differences between groups were found on the single primary health outcome initially included in the cost-effectiveness analysis. In contrast, relevant and significant differences were identified on three of the four additionally assessed health outcomes. Conclusions based on the results of multiple cost-effectiveness analyses and acceptability curves were strongly in favour of the experimental intervention when including these three additional instruments. QALY results did not show differences between groups.

Discussion: Selecting between outcome measures for cost-effectiveness analysis in the field of mental healthcare appears to be an arbitrary process, which may have considerable consequences for the results of economic studies and subsequent policy decisions. It was argued that inconsistent results across the selected primary health outcome and additionally assessed health outcomes should explicitly be presented to decision-makers. Until there is consensus on a preference-based instrument suited for mental healthcare, various QoL instruments could be applied instead of instruments aimed at specific aspects of health.

Implications for Health Policies: Decision-makers in the field of mental healthcare should be careful when interpreting results of economic studies that included single outcome measures aimed at a specific aspect of health. Due to current reservations on the use of QALYs in mental healthcare, QALY outcomes should be considered in the context of the results of additionally assessed health outcomes.
Introduction

Decision-makers in the field of healthcare are confronted with limited resources. As a consequence, priorities have to be set for health programs and interventions, both new and existing ones. Economic evaluation can provide relevant information to aid decision-makers in this complex process of prioritising and seems to be accepted as a useful instrument for rational policy decisions (1). In this context, it is essential that economic evaluations provide valid and reliable information on both costs and health outcomes of alternative interventions. For economic studies in the area of mental healthcare, however, the assessment of health outcomes is associated with complications and considerable controversy. Guidelines on the design of economic evaluations generally recommend the use of preference-based health outcomes, in particular Quality-Adjusted Life Years (QALYs: 2), which enable decision-makers to compare outcomes across disorders and studies. The QALY combines quantity with quality of life, where quantity is expressed in terms of life years gained from an intervention, and quality as the preference (utility or value) for health states. Although the QALY concept seems highly relevant from a decision-making perspective, there is no consensus on the use of QALYs in economic evaluation (3). In the area of mental healthcare, various specific concerns exist about QALYs. Unsatisfactory results have been found in several studies that tried to measure utilities in mental healthcare, and there are various difficulties related to the valuation of health states measured in people with severe mental illness (4). Moreover, physical aspects are often overrepresented in instruments that are used for deriving QALYs, which seems to make such instruments less relevant for measuring health in mental illness. Consequently, only few economic studies in mental healthcare actually applied QALYs. By far the most conducted economic evaluations were designed as cost-effectiveness analyses (5) that included single outcome measures aimed at a specific aspect of health (6). The single outcome measure included in cost-effectiveness analysis is assumed to best reflect relevant health consequences of interventions under study. However, selecting between available outcome measures in the field of mental healthcare can be complicated since mental illness, and particularly severe mental illness, often affects multiple domains of health and functioning (7-10). Moreover, the selection of a single outcome measure appears to be influenced by the perspective from which the patient is considered (11-13). Clinicians typically focus on psychiatric symptoms, relatives of the patient may stress the importance of social functioning, while unmet needs in the area of mental healthcare and rehabilitation may be most relevant for the patient. In practice, the primary outcome measure is usually selected by clinicians or clinically oriented
researchers, which often seems to involve rather arbitrarily choosing one relevant aspect of health over another. Unfortunately, the relevance of a selected primary health outcome for policy decisions (14) is not always taken into account. In the current paper, the potentially negative consequences of selecting between various outcome measures for economic evaluations in mental healthcare will be illustrated by the results of a previously conducted cost-effectiveness analysis in patients with severe mental illness.

Method

Design economic evaluation
Analyses presented in this paper were based on the data of a study focusing on the cost-effectiveness of the HIT (Hallucination focused Integrative Treatment) programme in patients with schizophrenia and a history of persistent auditory hallucinations (15). HIT integrates cognitive behaviour therapy with various additional forms of intervention, including psycho-education and single family treatment (16). In total 76 patients were randomly assigned to two treatment arms, HIT or care as usual (CAU). The HIT programme was provided during the first 9 months of the study and consisted of approximately 11 contacts with HIT therapists in addition to regular care. CAU was not standardised in the present study and mainly consisted of home visits, psychiatric and social management, and maintenance of medication. The economic evaluation was performed from a societal perspective, i.e. a wide range of costs in and outside the healthcare sector was registered. Medical costs that were assessed included costs of hospital admissions, medication use, and contacts with various healthcare professionals. Costs outside the healthcare sector included costs of informal care, travel costs, and costs related to productivity losses. The Positive And Negative Syndrome Scale (PANSS; 17) was selected as the primary outcome measure during the design phase of the study. The PANSS is a 30-item, semi-structured interview with the patient on psychiatric symptoms (including hallucinations). Power analyses were based on characteristics of the PANSS in the patient population under study. Health outcomes and costs of included patients were registered prospectively during a period of 18 months. Measurement took place at nine-month intervals, starting at the time of inclusion (T0, T9, T18). Since the economic evaluation was an integral part of a clinical study on HIT (18-20), several additional instruments were administered at the same time as the PANSS to assess a range of relevant clinical aspects. Outcomes that were assessed included social functioning,
subjective burden of auditory hallucinations, and Quality of Life (QoL). Experienced researchers were responsible for the interviews with patients and the administration of questionnaires. Results were based on the data of 83% (n=63) of the initially included patients for whom all the relevant costs and health outcomes could be assessed during the study. Longitudinal analyses were conducted in accordance with a complete case approach, no imputation techniques were applied.

**Instruments**

Main characteristics of the additionally administered instruments are briefly discussed to provide the necessary background information. The Groningen Social Disabilities Schedule (GSDS; 21) is a valid and reliable semi-structured observer-rated interview that addresses social functioning. Total scores of the 7-item version of the questionnaire can range from 0-21. The Auditory Hallucination Rating Scale (AHRS, 22) is a widely used semi-structured interview with satisfactory psychometric properties. The AHRS measures the subjective burden of auditory hallucinations, control over voices and interference with daily functions. Total scores of the AHRS vary between 0 and 32. The EuroQol instrument (23) is a self-administered QoL questionnaire commonly applied in economic studies, consisting of five items that address mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. EuroQol results were transformed into utilities by applying the algorithm of Dolan, which is based on the valuation of health states by the general population (24). Subsequently, QALYs were derived by combining utilities with life years. The World Health Organisation Quality of Life (BREF) assessment (WHOQoL-BREF; 25) is a well-validated 26-item version of the WHOQoL-100 instrument and measures four domains of QoL, i.e. physical, psychological, social, and environment. Outcomes of the abbreviated instrument can be transformed to a scale ranging from 0 to 100.

Lower scores on the PANSS, AHRS, and GSDS indicated a better health status. Accordingly, patients who improved during the study had lower scores on these instruments at the last follow-up (T18) compared to the baseline assessment (T0). In the current paper, change from baseline scores (T18 minus T0) calculated for these instruments were multiplied by -1 to represent an improvement in functioning by positive difference scores. This facilitated comparisons with other studies where better health status is typically represented by higher scores and positive change from baseline scores.

**Economic analyses and data analytic procedures**

Incremental cost-effectiveness ratios (ICERs) were calculated for each of the administered instruments, mean cost and effect differences were presented in cost-
effectiveness planes (26, 27). Furthermore, bootstrap simulations addressed uncertainty surrounding the calculated ICERs (28). Bootstrapping is an iterative method that consists of randomly selecting patient data (with replacement) from the observed population to create a simulated distribution of data. ICERs were calculated for each of the bootstrap iterations (5000 in the present study), and simulated values of the mean estimates for the cost and outcome differences were presented in cost-effectiveness planes. Cost-effectiveness acceptability curves (CEACs; 29, 30) were based on each of the assessed health outcomes. CEACs provide information on the probability that an intervention will be cost-effective, depending on what decision-makers are willing to pay for an additional unit of health outcome (monetary threshold value). Differences in health outcomes between groups were analysed by Student’s T-tests focusing on change from baseline scores.

Results

Cost-effectiveness analysis

In this section, a summary of the most relevant results of the cost-effectiveness study focusing on the HIT intervention is provided. For a detailed description the reader is referred to the original publication (15). Table 1 presents the outcomes on the PANSS and the total costs of both groups during the study.

Table 1. PANSS outcomes and costs (T0-T18)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>HIT (n=31)</th>
<th>CAU (n=32)</th>
<th>Significance of difference (95% CI)³</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td>PANSS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>57.1</td>
<td>60.2</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>51.1</td>
<td>57.3</td>
<td></td>
</tr>
<tr>
<td>Change from baseline¹</td>
<td>6.1</td>
<td>2.9</td>
<td>n.s. (-4.23, 10.55)</td>
</tr>
<tr>
<td>Costs²</td>
<td>Mean costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18,237</td>
<td>21,436</td>
<td>n.s. (-12,050, 6,637)</td>
</tr>
</tbody>
</table>

¹ Positive change from baseline scores indicate improved functioning.
² All costs are in US dollars (price level of 2000).
³ Student’s T-test for independent samples. Lower and upper limits of the 95% confidence interval of the difference in means are provided between brackets. For costs, the nonparametric 95% confidence interval was assessed with the bootstrap method. Limits of the confidence interval indicated a not statistically significant difference in costs between groups.
Outcomes on the PANSS were slightly in favour of the HIT group (3.2 difference in change from baseline scores on a score range of 30-210). However, these differences were neither clinically relevant nor statistically significant. The mean costs of patients in the HIT group ($18,237) were lower than the mean costs of patients who received CAU ($21,436). The value of the calculated ICER was – $936 per point improvement on the PANSS. This negative value indicated that patients in the HIT group generated fewer costs and had better results on the PANSS. In additional analyses, bootstrap simulations were conducted. The HIT intervention dominated CAU in approximately 60% of the cases. In Figure 1, the cost-effectiveness acceptability curve (CEAC) based on the outcomes of the performed bootstrap analysis is presented. Increasing monetary threshold values per additional unit of health outcome first led to a slightly increasing probability that the HIT intervention would be cost-effective, but this probability never exceeded the 85% level. For values over $2,000 the probability slowly decreased towards 79%, due to the location of the joint density in the north-east and south-west quadrants of the cost-effectiveness plane (30).

Based on the presented results, it was concluded that differences in costs and health outcomes were in favour of HIT and seemed relevant from a decision-makers perspective. A statistically significant cost-effectiveness advantage of HIT over CAU could not be demonstrated.

Figure 1. Cost-effectiveness acceptability curve for PANSS outcomes
Additional health outcomes and cost-effectiveness analyses

The results of the outcome measures that were assessed in addition to the PANSS are summarised in Table 2, for a detailed description the reader is referred to the published outcomes of the clinical study (18, 20).

Table 2. Results of the additionally assessed outcome measures (T0-T18)

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>HIT (n=31) Mean (SD)</th>
<th>CAU (n=32) Mean (SD)</th>
<th>Significance of difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>10.2 (4.2)</td>
<td>11.3 (4.8)</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>7.4 (4.1)</td>
<td>10.6 (4.2)</td>
<td></td>
</tr>
<tr>
<td>Change from baseline&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2.8 (3.2)</td>
<td>0.7 (3.9)</td>
<td>p&lt;.05 (0.35, 3.96)</td>
</tr>
<tr>
<td>WHOQoL-BREF&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>57.6 (8.0)</td>
<td>59.3 (9.2)</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>62.9 (9.7)</td>
<td>58.9 (12.0)</td>
<td></td>
</tr>
<tr>
<td>Change from baseline&lt;sup&gt;1&lt;/sup&gt;</td>
<td>5.4 (10.3)</td>
<td>-0.4 (10.9)</td>
<td>p&lt;.05 (0.29, 11.23)</td>
</tr>
<tr>
<td>AHRS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>24.7 (3.9)</td>
<td>23.5 (4.8)</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>14.2 (8.3)</td>
<td>19.2 (9.8)</td>
<td></td>
</tr>
<tr>
<td>Change from baseline&lt;sup&gt;1&lt;/sup&gt;</td>
<td>10.5 (9.1)</td>
<td>4.3 (9.0)</td>
<td>p&lt;.01 (1.58, 10.70)</td>
</tr>
<tr>
<td>QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0 – T9</td>
<td>0.5 (0.2)</td>
<td>0.5 (0.2)</td>
<td></td>
</tr>
<tr>
<td>T9 - T18</td>
<td>0.5 (0.2)</td>
<td>0.5 (0.2)</td>
<td></td>
</tr>
<tr>
<td>Total QALYs</td>
<td>1.0 (0.3)</td>
<td>1.0 (0.3)</td>
<td>n.s. (-0.15, 0.14)</td>
</tr>
</tbody>
</table>

<sup>1</sup> Positive change from baseline scores indicate improved functioning.

<sup>2</sup> n=30 in both groups.

<sup>3</sup> Student’s T-test for independent samples. Lower and upper limits of the 95% confidence interval of the difference in means are provided between brackets.

Differences between groups are presented in terms of ‘change from baseline scores’ on the additionally assessed outcomes, i.e. social functioning, QoL, and subjective burden of auditory hallucinations. Results of the GSDS showed that social functioning of patients in the HIT group improved significantly (p<.05) compared to the CAU group. On the WHOQoL-BREF, a statistically significant difference in terms of improvement in QoL (p<.05) was found in favour of the HIT group. Outcomes on the AHRS indicated that the level of distress decreased and the control over voices increased significantly in the HIT group compared to the CAU group (p<.01). However, QALY results (based on utilities derived from the EuroQol) did not show relevant differences between groups. Costs were already presented in the first part of the results in Table 1.
ICERs were calculated based on the outcomes of these four additional outcome measures, followed by bootstrap analyses (results are presented in Figure 2).

Figure 2. Results of economic analyses based on the GSDS, WHOQoL-BREF, AHRS and QALY outcomes

A. GSDS

B. WHOQoL-BREF

C. AHRS

D. QALY

For the GSDS, WHOQoL-BREF, and the AHRS, the HIT intervention dominated CAU in approximately 75% of the bootstrap replications. Virtually all the bootstrap outcomes were restricted to the right side of the plane, emphasising significant differences in health outcomes between groups. For the QALY results, bootstrap outcomes were more evenly distributed across the plane, which indicated that there were no relevant differences.

Figure 3 presents the CEACs calculated for the GSDS, WHOQoL-BREF, AHRS, and QALY outcomes. Please note that the four presented CEACs cannot directly be compared, because score ranges of the applied instruments vary. For instance, one (additional) unit of health outcome on a range of 0-20 (GSDS) should be
interpreted differently compared to one unit on a range of 0-100 (WHOQoL-BREF).

Figure 3. Cost-effectiveness acceptability curves

A. GSDS

B. WHOQoL-BREF

C. AHRS

D. QALY

Figure 3 shows that if a decision-maker is willing to pay $2,000 for an additional unit of health, the probability that HIT would be cost-effective tends towards 100% for both the WHOQoL-BREF and the AHRS. The CEAC for the GSDS demonstrated the same trend, but the monetary threshold value placed on an additional unit of health outcome would be slightly higher. The CEAC for the QALY results suggested that the probability of cost-effectiveness was indifferent to monetary threshold values up to $8,000 per QALY gained.
Discussion

Although it is strongly recommended to use preference-based outcome measures in economic evaluations, most studies in mental healthcare included outcome measures aimed at a specific aspect of health (6). The current paper focused on the difficulties associated with the selection of a single primary outcome measure in severe mental illness. These difficulties were illustrated by an economic evaluation that examined the HIT intervention in patients with schizophrenia and persistent auditory hallucinations. Results of the conducted cost-effectiveness analysis, with the PANSS as single primary outcome measure, were presented. Subsequently, an overview was provided of the results of the additionally assessed health outcomes, i.e. the GSDS (social functioning), WHOQoL-BREF (quality of life), AHRS (burden of hallucinations), and QALYs. Subsequently, ICERs were calculated, the bootstrap method was applied to assess uncertainty, and cost-effectiveness acceptability curves were estimated for all the measured health outcomes and costs. Results of the economic analyses based on the GSDS, WHOQoL-BREF, and the AHRS were in favour of HIT and seemed highly relevant from a decision-making perspective. In contrast, differences between groups were less obvious for economic analyses based on the PANSS, and nearly absent for QALY outcomes. The inconsistency of results across health outcomes illustrates the problems that may arise for economic evaluations in mental healthcare when selecting between different outcome measures. In the presented study, decision-makers would have received insufficient information if only the results of one of the outcome measures had been available. Depending on the health outcome available, policy decisions could have ranged from strongly favourable to indifferent towards the implementation of HIT in current healthcare systems. Therefore, it seems important for economic analysts to at least describe inconsistencies in (additionally) assessed health outcomes when reporting results to decision-makers. Moreover, including conflicting results in supplemental economic analyses may provide relevant additional information. In the literature, there are indications that the inclusion of various outcome measures in economic evaluation is considered to be more informative for decision-makers than strictly focusing on a single outcome measure as is traditionally done in cost-effectiveness analysis (31, 32). Awareness of inconsistencies in health outcomes seems to be highly relevant for the policy-making process. In some situations, decision-makers may eventually have to conclude that adequate policy decisions cannot be based on presented results, due to such inconsistencies.

The finding that QALY outcomes did not demonstrate clear differences between groups, in contrast to three of the other outcome measures, was difficult to
interpret. In the presented study, the quality component of the QALY consisted of utilities derived from the EuroQol questionnaire. Results showed that there were no clear differences between groups in terms of utilities. The quantity component of the QALY, life years gained, was not directly affected; none of the included patients died during the study period. A closer examination of the current results revealed that utilities based on the valuation of (EuroQoL) health states by the general public seem to be particularly insensitive for changes in health of patients with mental illness (19). Various concerns related to the use of QALYs in mental healthcare have already been expressed in the literature (4). Although the EuroQol is among the most widely used instruments for assessing QALYs in economic evaluations, there are no publications available that adequately examined the validity of this instrument (or derived QALYs) in the area of mental healthcare. The mere absence of such studies was recently also noted by Lewis and colleagues (33), who strongly recommended research on the validity of the EuroQoL in patients with severe mental illness. Current findings seem to further support their claim for adequate research on this topic. Clarity on the validity of this instrument in mental healthcare may contribute to changing the present reluctance of mental healthcare professionals to apply preference-based outcome measures.

Most published economic studies in the area of mental healthcare have focused on primary health outcomes aimed at specific aspects of health. Although such outcomes are relevant from a clinical point of view, they cannot always provide decision-makers with information that is useful for policy decisions. The overall well-being (or QoL) of an individual is considered to be a much more relevant indicator for the effectiveness of an intervention for decision-makers. There are various instruments available that can accurately assess QoL. In the presented study, the WHOQoL-BREF was one of the administered questionnaires, which is considered to be a reliable and valid generic QoL instrument, both in patients with somatic and mental illnesses (34). Furthermore, disease-specific QoL instruments have been developed for a wide range of mental illnesses and disorders. It should be noted that QoL instruments are only directly applicable in cost-effectiveness studies if results can be expressed in a single overall score. Reliable and valid QoL instruments fulfilling this requirement may currently be considered as a useful alternative for instruments addressing specific aspects of health in cost-effectiveness analyses in mental healthcare.

To conclude, by far the most economic evaluations in mental healthcare were designed as cost-effectiveness studies focusing on single primary health outcomes aimed at a specific aspect of health, instead of preference-based outcomes recommended by guidelines. Selecting between single primary health outcomes for cost-effectiveness analysis may have considerable consequences for the results of
economic evaluations and subsequent policy decisions. Inconsistency of results across the selected primary health outcome and additionally assessed clinical health outcomes should at least be presented to decision-makers, since awareness of this inconsistency could be highly relevant for the decision-making process. For the moment, researchers or clinicians planning to conduct economic studies in mental healthcare may consider to use QoL instruments instead of instruments focusing on specific aspects of health. Since there are advantages related to the use of preference-based outcomes in economic evaluations, it is important to adequately assess the validity of instruments like the EuroQoL and eventually reach consensus on the use of preference-based outcomes in mental healthcare.
References


32. Hoffmann C, Stoykova BA, Nixon J, Glanville JM, Misso K, Drummond MF. Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. Value Health 2002; 5: 71-78.

