The Stuttgart-Heidelberg Model of Active Feedback Driven Quality Management: Means for the Optimization of Psychotherapy Provision

Hans Kordy and Stephanie Bauer
(Center for Psychotherapy Research Stuttgart, Germany)

(Received January 3, 2003 / Accepted April 30, 2003)

ABSTRACT. Quality management deals with the evaluation of psychotherapeutic treatment. A central aspect concerns the development of appropriate assessment batteries and evaluation criteria. The Stuttgart-Heidelberg (S-H) model represents a system providing concepts, psychometric tools and a computer software, developed for the purpose of active feedback driven quality management. The key information of the Stuttgart-Heidelberg Model is on the outcome of the individual treatment. The assumption is that psychotherapeutic care can be improved by providing information on shortcomings of the delivered care to practitioners, because problem solving processes will be stimulated by the feedback. The present contribution introduces the assessment inventory, the standardized outcome evaluation and the different feedback tools of the S-H model. A systematic study including 1715 patients from a psychosomatic hospital documents the validity of the approach. The empirical findings encourage the call for transparency about what happens in clinical routine – i.e. about the applied treatments, their results and their costs. Implications for further optimization of health care provision are discussed.

RESUMEN. La gestión de la calidad busca la evaluación del tratamiento psicoterapéutico. Un aspecto central se relaciona con el desarrollo de baterías de evaluación y criterios de evaluación adecuados. El modelo Stuttgart-Heidelberg (S-H) representa un sistema que proporciona conceptos, instrumentos psicométricos y un programa informático desarrollado para la gestión de la calidad basada en el feedback activo. La información central del modelo Stuttgart-Heidelberg es el resultado individual del tratamiento. El planteamiento es que la psicoterapia puede mejorarse si proporcionamos información sobre los resultados terapéuticos (en especial los negativos), pues los procesos de solución de problemas se estimularán por el feedback recibido. El presente trabajo presenta un inventario de evaluación, la evaluación estandarizada de los resultados y las diversas herramientas de feedback del modelo S-H. Un estudio sistemático incluyendo 1715 pacientes de un hospital especializado en trastornos psicosomáticos documenta la validez de este abordaje. Los resultados empíricos refuerzan una estrategia de transparencia acerca de lo que acontece en la práctica clínica – por ejemplo, acerca de los tratamientos administrados, sus resultados y costos. Implicaciones para la posterior optimización de los servicios de salud son discutidos.


RESUMO. A gestão da qualidade lida com a avaliação do tratamento psicoterapêutico. Um aspecto central relaciona-se com o desenvolvimento de baterias de avaliação e critérios de avaliação adequados. O modelo Stuttgart-Heidelberg (S-H) representa um sistema que fornece conceitos, instrumentos psicométricos e um programa de computador desenvolvido para a gestão da qualidade baseada no feedback activo. A informação central do Modelo Stuttgart-Heidelberg é o resultado individual do tratamento. O pressuposto é o de que a psicoterapia pode ser melhorada se fornecermos informação acerca dos resultados terapêuticos, especialmente os negativos, pois os processos de solução de problema serão estimulados pelo feedback recebido. A presente contribuição apresenta um inventário de avaliação, a avaliação estandardizada dos resultados e as diversas ferramentas de feedback do modelo S-H. Um estudo sistemático incluindo 1715 pacientes de um hospital especializado em doenças psicosomáticas documenta a validade da abordagem. Os resultados empíricos encorajam uma estratégia de transparência acerca do que acontece nas rotinas clínicas – i.e. acerca dos tratamentos administrados, os seus resultados e custos. Implicações para posterior optimização dos serviços de cuidados de saúde são discutidos.


Introduction: Quality management in psychotherapy provision
Psychotherapy looks back on a successful history - and forward to an exciting future. Close co-operation between researchers and clinicians – often combining these orientations in person – made this successful development possible. Clinicians challenged researchers by clinical questions and research findings stimulated clinicians to try out new things and to raise new questions. Numerous studies demonstrated the efficacy of psychotherapy
Figure 1 describes this development from a general perspective (Sachverständigenrat für die Konzertierte Aktion, 1995). The innovation phase came to an end during the eighties. This was followed by the diffusion phase, i.e., by extension of service provision for various purposes by an increasing number of providers. The current phase is still characterized by diffusion, but more and more saturation has to be taken into consideration. As explained by the schema evaluation gets an important steering role for the transition from the innovation to the diffusion phase in deciding whether an innovative service alternative allows to achieve the given objectives with reasonable effort. The great number of evaluation studies of psychotherapy during the seventies and eighties corresponds with this (e.g. Kordy and Kächele 1997; Strupp and Howard, 1992). In approaching the level of saturation the attention shifts to quality assurance. Correspondingly the number of

![FIGURE 1. Evaluation and quality management in the innovation process.](image-url)
publications on developing and testing models of quality assurance increased during the nineties. Quality management and outcome orientation are considered as efficient tools for the optimization of service provision. The expectation is that full transparency about what happens in everyday clinical practice will open up new ways of systematic learning for all who can contribute to the optimization, i.e. the therapists (providers) and the patients (consumers) as well as health care managers, health insurance managers, and health politicians.

The Center for Psychotherapy Research in Stuttgart started an initiative in the early nineties. Together with clinical partners from the Psychiatric Institute of the University of Heidelberg the Stuttgart-Heidelberg Model of Active Internal Quality Management was developed (Kordy 1992; Kordy and Lutz, 1995; Kordy, Hannöver, and Richard, 2001; Lutz, Stammer, Leeb, Dötsch, Bölle, and Kordy, 1996). Currently it is applied in more than 20 German hospitals specialized on psychotherapeutic and psychosomatic medicine (and two Swiss hospitals) and treatments of about 5000 patients are monitored per year. Although it has been developed for service provision under German conditions – e.g. with much larger proportion of inpatient psychotherapy than in other countries – the principles may apply also in other countries. The current article tries to follow the guidelines proposed by Bobenrieth (2002).

The Stuttgart-Heidelberg Model

Every failure is a treasure! This motto of quality assurance (e.g. Berwick, 1989) guides the Stuttgart-Heidelberg Model. The model is characterized as an active, internal approach that puts information feedback and continuous learning in the center. The key assumption is that providing information on shortcomings of the delivered care to therapists and patients will stimulate problem solving processes and ultimately lead to improvement. For this purpose the model provides four essential components: (1) an inventory for the collection of the relevant data; (2) a standardized rule for the evaluation of treatment outcome which enables the detection of shortcomings; (3) feedback tools that transport the message to those who can make use of it, and (4) the concept of clinical quality circles as a communication and learning culture (Selbmann, 1996).

The inventory

The inventory should be sufficient, but also parsimonious, to answer the four key facets of quality assurance, i.e. who is served by what means at what costs and with what results? The Stuttgart-Heidelberg Model is very flexible with respect to the measures. Instead of constructing own psychometric scales – and so perhaps save the license fee for established measures - it takes advantage of the existing, sufficiently standardized and widely used measures. Although it allows the users to use their own measures a set of core measures is strongly recommended. This ensures external comparisons between providers what is not only of clinical or scientific interest, but also required by the respective law, the Health Reformation Law 1988 (Deutscher Bundestag, 1988). Assessments are taken from the patients’ as well as from the therapists’ view. The three main dimensions of psychic, of bodily and of interpersonal/social impairment are taken into account.
The core set comprises (a) patient view: the Symptom-Checklist (SCL-90 R) (Franke, 1995), a list of 24 bodily complaints (the Giessener Beschwerdebogen, GBB) (Brähler and Scheer, 1995), the 64-item version of the Inventory of Interpersonal Problems (IIP) (Horowitz, Straus, and Kordy, 1994), a Life Satisfaction Questionnaire (the Freiburger Bogen zur Lebenszufriedenheit, FLZ) (Fahrenberg, Myrtek, Wilk, and Kreutel, 1986) and finally a retrospective assessment of health status changes; (b) therapist view: the clinical assessment of impairment of functioning (Beeinträchtigungsschwere-Score, BSS) (Schepank, 1987) and a retrospective assessment of health status changes like that done by the patients. Assessments are undertaken at beginning and ending of the treatment, assessments at follow-up or intermediate time points are recommended.

The standardized outcome evaluation

What causes the main concerns of the psychotherapy providers is transparency of the outcome of psychotherapy. One at first is confronted with the failure before one can turn it into a treasure. This is not only a psychological problem – that it is obviously – it bears also economic risks because of the competitive situation in the health service market. This mix of real and fantasized risks makes a general consensus definition of outcome standards not very probable. The Stuttgart-Heidelberg Model gave therefore priority to a pragmatic strategy of small steps and developed the concept of signal cases that allows detection of cases which outcomes call for a second view together with peers. So the signal case is a stimulus to initiate problem solving processes and not a final judgement by a formal authority that clinicians have to surrender to. The emphasis on that one will learn how and what to learn from alarm signal cases usually eases an agreement on the basic principles of the standardized evaluation and the pragmatic operationalization. Three principles have turned out to be agreeable – at least for the users of the model (and the market will decide whether this model finds sufficient number of users): (P1) patient is still in strong need for treatment, (P2) there are indicators of vital threat for the patient, (P3a) the patient did not achieve sufficient positive changes or (P3b) even left treatment in a worse status than he or she entered.
FIGURE 2. Evaluation algorithm of the Stuttgart-Heidelberg Model.
These principles are then operationalized according to the special conditions for the service provision of the users. Figure 2 describes the standard evaluation for inpatient psychotherapy in Germany. Most of the users of the Stuttgart-Heidelberg Model are hospitals which serve the broad mix of patients with neurotic, psychosomatic or mental disorders. For these patients and this treatment setting it is suggested to consider patients (1) as in need for inpatient treatment if their functional impairment is worse than that of 68% of the usual patients mix in such a specialized hospital at admission, i.e. shows a BSS>8; (2) as under vital threat if he or she has strong or extreme suicidal thoughts when leaving the hospital; (3a) as not sufficiently improved if one finds more scales with reliable negative change than with reliable positive change (for the concept of clinically meaningful reliable change see e.g., Jacobson and Truax, 1991; Kordy and Senf 1985) or as ending treatment in a worse state if that is indicated by items evaluating change directly like change of general functioning evaluated by the therapist (THE) or patient (PAE, item 3). Further details are described in Figure 2. Patients who meet one of these criteria at discharge are considered as signal cases and their treatment will be clinically reflected in a team meeting which is called clinical quality circle in this context. The operationalization may be different for different settings, e.g. outpatient treatment (Scheidt et al., 1998) or specific patient groups, e.g. eating disorders (Kordy, Richard, Herrman, Murphy, Treasure, and Charpentier, 1999).

The feedback tools

As mentioned above the Stuttgart-Heidelberg Model puts emphasis on learning from failure and thus problem detection is considered as initial step for problem solving processes. To give that an actual chance in everyday routine the information flow has to be efficient. That means, that meaningful information in parsimonious format has to be provided in time to those who can make use of it (c.f. Lugon and Secker-Walker, 1999), i.e., in inpatient treatment settings with priority to the clinical team. A specific computer software, the AKQUASI, facilitates this. It provides information feedback in various ways. Among these, two tools are of especial importance for the management of the quality of the provided psychotherapeutic service.

The standardized outcome evaluation form

The key information of the Stuttgart-Heidelberg Model is on the outcome of the individual treatment. The standardized outcome evaluation form summarizes the information relevant for the outcome on two pages. An example is given by Figures 3 and 4. It is recommended to read clock wisely beginning with the global information in the middle of the right side. In this example, a patient of 46 years with diagnose of an adjustment and personality disorder (F43, F60 according to ICD-10) was treated for 23 days in a hospital specialized on psychotherapeutic medicine and psychosomatics. This patient was identified as a signal case because he showed more scales with reliable negative change than scales with positive change. The graphical display in the right lower corner gives an overview on the changes: a + (-) represents a reliable positive (negative) change while ++ (—) represents a reliable change into (out of) pre-defined normal range. In the left lower corner the legends for the scales are listed. Detailed information on the
assessments for this patients is given by the table on the top. The first row reports the intake assessments, while the second describes the discharge status of the patient. In the 3rd and 4th row one can find reference or cutoff-points which rely on norms or non-patients samples. In this example, the patient describes himself as moderately to highly impaired at beginning of treatment. High scores on somatic complaints (GBB), social uncertainty (SCL-90, scale 3), paranoic thinking (SCL-90, scale 8) and several IIP-scales call for attention. In contrast to this, the therapists sees only mild functional impairment (BSS). At discharge the therapist came to a similar assessment as at admission, while the patient reported many changes to the worse, several of them are substantial. The negative view of the patient of his treatment outcome corresponded to the low score for the therapeutic relationship and satisfaction with treatment at discharge.

In the clinical discussion the therapist explained that the treatment ended prematurely, because both, patient and therapist, saw no chance for substantial progress. The therapist had assessed the therapeutic alliance as low at beginning of the treatment already, he had especially mentioned doubts on the motivation of the patient and whether they shared similar goals for this treatment. Thus the therapist understood the patient’s report on the changes to the worse as reaction to the subjectively disappointing and discouraging treatment and its end. Nevertheless, the functional impairment did not call for revision of the assessment from therapist’s point of view. The clinical team agreed with the opinion of the therapist that the patient did not really get involved in his treatment. He obviously had expectations that could not be met by the hospital. Under this perspective the team supported the decision to stop the treatment prematurely. Although there had been hints for a negative prognosis from beginning on the team considered it justified having tried the treatment. Nevertheless, there was broad agreement that the actual poor treatment outcome gives reason for the alarm signal.

The quality control chart

The quality control chart (QC-Chart) allows the monitoring of the rate of alarm signal cases over time and therefore supports the service provision in a clinical unit. Figure 5 gives an example for a 110 bed hospital specialized on psychotherapeutic medicine and psychosomatics. The average length of stay in this hospital is 33 days. Usually a QC-Chart shows the failure rate over time, where the failure rate is estimated by a random sample drawn from the total set of observation in a certain time period, e.g. on a weekly base. The limited number of patients in hospitals of this size per year suggests modification of this procedure.
Thus, the Figure 5 represents a descriptive QC-Chart where the failure rate is calculated for every 30 cases (the dates on the x-axis give the admission date of the first case of the 30). Two horizontal lines facilitate the interpretation of the chart. The baseline ($p_0$) serves as standard or benchmark. It is obtained from the observed failure rates in this hospital from the period 1995/96. The second (broken) line represents the so called action limit. It is similarly calculated as a confidence limit (with an alpha of 10% and n=30 in this case). A failure rate beyond the action limit calls for action, i.e. it is interpreted as an alarm signal for a possible deterioration of the outcome quality of this unit. In such a case, the team would be called together to search for possible reasons and if necessary for counter actions. The illustrated example shows a decrease of the failure rate from summer 1999 to summer 2000 followed by a phase of increasing failure rates. Discussion with the clinical team suggested that this increase was related to changes in the team, two psychotherapists and one nurse had left by the end of 2000 and were replaced then. Although it is open whether these fluctuations in the clinical team actually caused the increase of failure rates the team found it stimulating to reflect this possibility. Fortunately, the failure rate went down to the former low level in winter 2001 (what perhaps supports the speculation).
The validity of the approach

The Stuttgart-Heidelberg Model is considered as helpful by providers and as relevant by consumers of psychotherapy. Several hospitals have used the system for years now, and do not plan to stop that. In a systematic survey in one of the hospitals (including 1715 patients) the vast majority of the patients rated such a routine quality assessment as very or rather important (92.6 %). This held true for alarm signal cases (92.9 %) as well as for cases with good or very good outcomes (92.4%). Perhaps even more convincing are the high response rates in the follow-up investigations routinely conducted by that hospital. About 80.7% (81.5% of the non-signals, 75.6% of the signals) of the former patients invested about 30 minutes to fill out the questionnaires 6 months after leaving the hospital. Of course, that does not make further investigations of the usefulness of the approach and the validity of its central concept, the alarm signal case, unnecessary. A detailed exploration of the follow-up data of 631 cases by classification and regression tree methodology demonstrated that the alarm signal has substantial predictive value. Cases assessed as signal cases at end of treatment showed a risk of 59.5 % for being evaluated as signal case at 6 months follow-up which is about 3 times the risk of non-signal cases at discharge (22.4 %). This held true also when relevant co-variates were taken into account (Kordy et al., 2001).

FIGURE 6. Source of information contributing to signals.

In the above mentioned survey 340 patients (20.0 %) were assessed as alarm signal cases at discharge. The Venn diagram in Figure 6 explains which source of information these assessments rely on. Obviously the data provided by the therapists are not the most
important. Only 9 signal cases (2.6 %) were detected by the therapists alone, while more three quarters (78.8 %) rely on information from the scales of the patient report alone. These numbers do not necessarily indicate disagreement between therapists’ view and the psychometric measures of patient self-report. At least when confronted with the result of the standardized evaluation therapists usually agree. Therapists were routinely invited to compare their clinical view with the standardized assessments for their patients (n = 1715). For about half of the cases (55.4 %) therapists used this opportunity to comment from clinical perspective; there was no substantial difference between alarm signal cases (56.2 %) and non-signal cases (52.0 %).

**TABLE 1.** Concordance rates between standardized evaluation and therapist’s view.

<table>
<thead>
<tr>
<th></th>
<th>Non signal cases</th>
<th>Signal cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Confidence limits</td>
<td>Confidence limits</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Total sample (n = 954)</td>
<td>708</td>
<td>91.0</td>
</tr>
<tr>
<td>F32: depressive episode (n = 31)</td>
<td>93</td>
<td>98.1</td>
</tr>
<tr>
<td>F33: recurrent depressive episode (n = 23)</td>
<td>40</td>
<td>96.2</td>
</tr>
<tr>
<td>F34: persistent affective disorder (n = 48)</td>
<td>40</td>
<td>100.0</td>
</tr>
<tr>
<td>F31: anxiety disorder (n = 42)</td>
<td>52</td>
<td>97.0</td>
</tr>
<tr>
<td>F43: adjustment disorder (n = 275)</td>
<td>123</td>
<td>96.0</td>
</tr>
<tr>
<td>F45: substance disorder (n = 130)</td>
<td>66</td>
<td>98.4</td>
</tr>
<tr>
<td>F40: obsessive disorder (n = 68)</td>
<td>55</td>
<td>97.0</td>
</tr>
<tr>
<td>F50: eating disorder (n = 53)</td>
<td>31</td>
<td>96.1</td>
</tr>
<tr>
<td>F90: personality disorder (n = 193)</td>
<td>70</td>
<td>97.4</td>
</tr>
<tr>
<td>Other (n = 77)</td>
<td>53</td>
<td>98.0</td>
</tr>
</tbody>
</table>

RIPCS/IJCHP, Vol. 3, Nº 3
In 86.5% of the cases the standardized evaluation was in concordance with the clinical impression (Table 1). This high rate of concordance was confirmed in the various diagnostic subsets although slightly different rates of signal cases were observed there. If one looks at the agreement for signals (57.3%) and non-signals (94.3%) separately, one sees substantial differences at first glance. Closer inspection of the explanations given by the therapists shows that actual dis-concordance is rare also for the signal cases: in only 1 out of 5 of those dis-concordant signal cases the therapists actually evaluated the outcome as positive by clinical impression (8.5% of all signal cases), while in all other cases the negative result is attributed to treatment external reasons.

Discussion

Health politics challenged the field by claiming quality assurance. Increased transparency about what happened in clinical routine – i.e., the applied treatments, their costs and their results – should help to control the costs and assure the quality of health services. All WHO member states agreed on the development and implementation of appropriate measures. In Germany, for example, the parliament committed service providers and health insurers by laws (e.g. Deutscher Bundestag, 1988) to look for an agreeable approach. The Stuttgart-Heidelberg Model is responding to this. It provides concepts, psychometric tools and computer software which allow to turn the external initiative into an internal management task. More than 20 hospitals specialized on psychotherapeutic medicine and psychosomatics have implemented it, the continuing use for more than 6 years proves feasibility in clinical everyday practice. The routinely collected data demonstrate the validity of the concepts and confirm the expectation that such an approach opens up systematic learning possibilities. The model puts emphasis on problem detection, not as goal but as mean of the optimization. Thus problem solving is given priority to sanctioning “black sheep”. However, it is not an easy task to confront oneself with failures and shortcomings. Uncertainties and mis-interpretations make it unnecessarily harder for clinicians and researcher to get involved in the problem orientation. For example, it is often overseen that the focus of the optimization process is rather on the provider, the providing unit, or the provision system than on the individual patient in treatment. Trivial, but also not always taken into account is the difference between optimal and maximal treatment. While patients in acute need and their therapist often tend to look for maximal help, from service point of view one has to keep a balance between the needs of a population and the possibilities and preferences of a society. That means, it is not the failure of an individual treatment that counts, but the failure rate sets the standards or benchmark for e.g. the hospital, the clinic, the private practitioner. Abstracting from the individual failure sounds cynical in a situation where a therapist is trying his or her best to help an individual subject, where empathy is required to help a suffering person. However, especially in the case of direct confrontation with failing there is a considerable risk to attribute that to therapist’s or patient’s characteristics or behaviors and thus to induce guilt feelings instead of motivation for systematic learning. This underscores the importance of a learning and communicative culture like the clinical quality circles in the Stuttgart-Heidelberg Model. There the clinical team meets with experts in quality.
assurance and they bring the feedback on the standardized evaluation of treatment outcome of an individual patient together with the clinical perspective as well as with the research perspective. Thus, shifting the attention to the failure rate creates distance to the individual failure and that supports the willingness for continuous learning.

Unfortunately, psychotherapy research does not provide answers to all questions that arise from quality assurance. Although psychotherapy has been proven efficacious for many disorders and under various conditions, there are inherent limitations. It is a matter of fact, that even the most efficacious treatment does not help every patient. Effect sizes of up to 1.5 are impressive, but although they tell that the bigger proportion of patients does better in psychotherapy than without, there is a more or less considerable proportion of patients who do not better than non treated. The situation is even more sobering if one looks at recovery rates and not only on improvement rates. In the best case, research tells what treatment gives the best chances to what patients - and that is a great advantage, of course. But the therapist has to find out whether that particular treatment works for the individual patient in front of him or her. More and more it is recognized, that the scientifically validated knowledge is only one component of optimal service provision, although a very substantial one. There are hints for a possible gap between what can be achieved under research conditions and what is actually achieved in clinical everyday practice. Valid information on the effectiveness of psychotherapy in clinical routine is merely not available, even less is known about the conditions and costs for improving the effectiveness by using new scientific knowledge. Not every innovation can be transferred to everyday practice, not every deserves it. For example, there is no sound empirical base to estimate what gain could be expected from the introduction of a new treatment intervention for what patients, even if that has demonstrated superiority in an efficacy study. Technological developments to facilitate such adaptations e.g. by outcome monitoring or outcome management are only at the beginning (e.g., Lyons, Howard, O’Mahoney, and Lish, 1997). Service system research, which takes the patient as consumer into account are exceptions as well as co-operation with payers. However, the optimization of health care provision requires co-operation and readiness for compromises between different partners with different interests, competencies, and responsibilities.

These are challenges for all established segments of health care provision, and psychotherapy is no longer in a protected niche. As diverse as the reasons are to seek psychotherapy as diverse the psychotherapeutic means. Psychotherapy provision comprises specific treatments for specific mental disorders as well as therapeutic approaches which are used as adjunctive or supplementing treatments for medical treatments, for example. Corresponding to the various reasons for psychotherapy the objectives differ. A patient who knows what he or she is looking for, has good chances to find appropriate help. That is the advantage of the broad offer. The disadvantage is that not everybody has easy access to the information needed for a good choice. Better any than no psychotherapy! However, various service providers compete for the attention of the patients, each of them claims to help more effectively, more appropriately, more comprehensively, more lasting or more cost efficiently than others. Therapists and patients as well as health care managers and politicians would definitely benefit from transparency. Knowing what one can actually expect from a provider would enable informed choices.
Heidelberg Model has proven to be useful for that. The positive experience with it in the clinical routine should encourage to adapt it to other service provision conditions or try out own approaches.

References


