Routine outcome monitoring & learning organizations in substance abuse treatment
Oudejans, S.C.C.

Citation for published version (APA):
CHAPTER 7

GENERAL DISCUSSION
Introduction

The main question in this thesis is whether it is possible to implement a type of ROM in SATCs that is feasible, generates valid data, is supported by key persons in the services, can be used to generate conclusions about effectiveness and, finally, contributes to the learning capacity of employees and the treatment system as a whole.

In this final chapter, the main findings regarding these research questions are summarized, discussed, and placed within the context of recent scientific developments in the area of ROM and quality in health care. Furthermore, methodological considerations that are important for the interpretation of the results are addressed.

For each research question, findings will be summarized and discussed. Next, final conclusions will be drawn. Finally, we will make suggestions for future research and summarize practice implications.

Feasibility of ROM and validity of outcome data

The first research question was whether it is possible to implement a type of ROM that is feasible and generates valid data. In order to collect routine outcome data, we implemented a call center that carried out telephonic follow-up interviews. This strategy curbed the time-consuming task of filling out questionnaires, a task that is often handed to treatment staff (Sperry et al., 1996). The data collected by the call center was used to feed back aggregated results over groups of patients afterwards, to formulate improvement projects for treatment programs, and to serve as accountability information to stakeholders.

The results of the study show that it is possible to implement a system for routine follow-up interviews in a feasible, low-budget manner. This is positive since it is common that selecting and tracking patients for follow-up interviews is either very time consuming and costly, is not successful, or it is unclear which patients are included resulting in dubious representativeness (Gerstein & Johnson, 2000; Harrison & Asche, 2001; McKenzie & Marks, 2003; Teruya et al., 2006). On the other hand, the employed method of telephonic follow-up interviews produced data that were sensitive to selection bias, limiting a valid interpretation of the results for the total treated population. Imputation procedures to correct missing data might be available and applicable, but more effort should be made to improve response rates.

The assessment of internal validity of the follow-up data in this study was limited to the examination of selection bias. Future studies on the validity of telephonic follow-up interviews should also include other aspects of internal
validity, such as interviewer variability and protocol integrity. In addition, we were restricted to the measurement instrument currently in use at the SATC where the study was performed. Fortunately, new instruments are being developed (Schippers et al., 2007), allowing more detailed outcome assessments similar to those in studies like MATCH and COMBINE (Anton et al., 2006; Project Match Research Group, 1997).

Another step to enhance the validity of the follow-up data is making sure administrative systems are up-to-date, since failure of patient tracking was the main reason for non-response. Most treatment centers use electronic patient records (EPRs) and no good facilities seem to exist in the EPRs to create frequent updates of contact information. Since the introduction of EPRs in the late 90’s, they were perceived as an unlimited source of data for researchers and policymakers. Now, a decade later, this promise has not been met. Data are often not entered or placed where information analysts do not look for them. This was also true for the particular case of retrieving contact details of patients in our validation study in Chapter 2. As a result, it was not possible to generate up-to-date lists of patients’ contact details. Implementing suitable databases and training users should be the first priority for using EPRs for ROM.

In Chapter 2 we discuss a different strategy to collect valid follow-up data, using a random sample of the treated population instead of trying to reach all treated patients, and spend all available resources on better follow-up rates. The detailed follow-up protocol formulated by Scott could be employed for this purpose (Scott, 2004). This procedure describes how contact details are kept up-to-date by employing an Engagement, Verification, Maintenance and Confirmation (EVMC) protocol. Using this protocol, Scott et al. were able to yield follow-up rates of 90% and more. This protocol is suitable to track patients in substance abuse treatment, whose lifestyle is often chaotic. The four stages of the EVMC protocol should be accomplished a few weeks before the actual follow-up date. It should be noted, however, that the repeated contacts involved in this protocol introduce a different threat to validity, because they might function as an intervention in itself and can result in altered reporting of substance use (Hansten et al., 2000; Scott, 2004).

The study in Chapter 2 is on the validity of aggregated data, where selection bias can hamper valid conclusions about the treated population. However, low response rates also hamper the quality of individual, concurrent ROM, since missing a substantial proportion of the treated population for outcome monitoring, means that only a small proportion of patients will profit from care adjusted to their specific needs.

Follow-up interviews can be employed as individual or concurrent ROM strategies and are alternatively referred to as “Concurrent Recovery Monitoring”,

CHAPTER 7 / GENERAL DISCUSSION / PAGE 102
“Extended Case Monitoring” or “Recovery Management Check-Ups”. They are presented as effective ways to reduce relapse and to recognize patients who relapse after treatment (Dennis et al., 2003; Scott et al., 2005; Stout et al., 1999). Tracking patients intensively can be desirable in this case, since the tracking itself serves as an intervention and enhances quality of care.

Support for ROM from key persons in SATCs

Since support for ROM amongst treatment professionals and lower management staff can be problematic, we also examined the facilitating and impeding factors for such a system in our project. We wanted to know if treatment staff supported the ROM system and which facilitating and impeding factors were experienced. The study in Chapter 3 describes these factors and the general support for a ROM project in SATCs.

Most treatment professionals and managers expressed their appreciation and the feedback sessions were considered as relevant and were valued positively. This implies that one of the main requirements for learning from (benchmarked) outcomes information was met (Mulder & de Loor, 2005). An important limitation is that we concentrated on one specific ROM strategy: Feedback of aggregated data. Caution is needed in the extrapolation of our findings from aggregated feedback to the situation of concurrent feedback. However, there are examples of data used for concurrent feedback that are also used for the aggregated feedback (de Beurs & Zitman, 2007; Zwanepol & De Groot, 2008), so our results are valid for this application.

Our findings are consistent with findings from other quality improvement projects, where internal and external support are emphasized, and are even operationalized as an instrument to assess the presence of conditions for successful quality improvement projects (Duckers, Wagner, & Groenewegen, 2008). But, as is known from social psychology theories (Eiser & van der Pligt, 1988), a positive attitude or expressing support is in itself no assurance for the actual behavior to take place, i.e. taking actions based on the feedback of outcome data. This was underlined by our own results. The proportion of treatment professionals that attended feedback sessions was modest at best and there was little or no evidence of the initiation of improvement projects based on the outcomes reported. One explanation is that attitudes and organizational culture can be obstacles in the steps between supporting quality improvement by mouth and actual deeds.

Individual clinicians in the treatment of individual patients take key decisions that affect outcome. Therefore, real improvement depends on the influence of our actions on these clinicians. As Teasdale formulated “[...]

CHAPTER 7 / GENERAL DISCUSSION / PAGE 103
unfortunately, recent surveys have shown that there is widespread disengagement and disenchantment among doctors and clinicians in all healthcare professions, particularly from what are perceived as political or managerial initiatives” (Teasdale, 2008). Similar impeding factors, such as the lack of staff buy-in, “data collection fatigue”, the burdensome data collection procedures, and misperceptions of the reasons for the study, were reported on in other research projects as well (Eagar, Trauer, & Mellsop, 2005; Harrison & Asche, 2001; Teruya et al., 2006).

Our study contributes to a better understanding of the processes involved in the implementation of improvement projects, and findings were corroborated by results from studies in similar projects. The next step is to build on the findings of this study, through quality departments that support teams in setting up ROM and translating results into improvements in routine practice.

**Effectiveness of treatment interventions in routine practice**

Another research question was whether it is possible to retrieve information from databases with routinely collected data to perform research on the effectiveness of treatments in routine practice. In the study reported in Chapter 4, we examined the feasibility and effectiveness of two evidence-based outpatient psychotherapeutic interventions for patients with alcohol use disorders. We concluded that the results of alcohol treatment RCTs are preserved in a routine practice, although the proportion of relapsing patients is still rather high.

One important limitation has to be noted when comparing the current naturalistic study in routine practice to existing data from RCTs. In the naturalistic study, we reached a follow-up rate of only 52% as opposed to the 80% regularly recommended as a lower limit. Since non-respondents were over-represented on variables that were negatively related to treatment outcome, our treatment outcomes are likely to be an over-estimation of the real treatment success, and this pattern was confirmed in the feasibility and validity study in Chapter 2. Therefore, conclusions and comparisons from this study should be made cautiously.

The question whether follow-up data collected for ROM can be used to generate conclusions about effectiveness of routine practice can be answered positively. It has been suggested before that the methodology of studies in the field of quality of care lack precision in operationalization and standardization (Nabitz, 2006). However, our study shows that it is possible to implement a project in which clear sampling procedures, standardized instruments, and appropriate statistical procedures are employed and in which conclusions regarding the effectiveness of evidence-based interventions can be drawn. In
pharmacotherapy research, these “phase IV” studies are part of the development process. However, in psychotherapy research this is less common. In the field of alcohol dependence, several extensive naturalistic outcome studies like DATOS, CALTOP and NTORS are available (Evans & Hser, 2004; Gossop et al., 2003; R. L. Hubbard et al., 2003; R.L. Hubbard, Craddock, Flynn, Anderson, & Etheridge, 1997), but for alcohol use disorders such studies are sparse. Some naturalistic pharmacotherapy studies are available, like acamprosate studies on pharmacotherapy and psychosocial interventions (Pelc et al., 2002; Soyka et al., 2002). The few available naturalistic studies on outcomes of psychotherapeutic interventions for alcohol use disorders, focus on intensive, long-term interventions consisting of a mixture of psychotherapeutic methods (Bottlender & Soyka, 2005; Gual et al., 1999; Ojehagen et al., 1994; Ojehagen et al., 1988). However, it is unclear whether these treatments were evidence-based. The current study is a contribution to the field because it evaluates the outcome of routine practice of an evidence-based treatment, and it is possible to compare results of daily routine practice to those of RCTs on psychotherapeutic interventions that were used in the current naturalistic study.

We concluded that the percentage of patients who reached abstinence in our study were comparable to those in RCTs on motivational enhancement therapy and cognitive behavioral coping skills therapy in project MATCH (Project Match Research Group, 1997): 25% versus 23% and 21% respectively. However, it must be taken into account that in our study a different follow-up window and a lower response rate possibly resulted in an over-estimation of the abstinence rate compared to the MATCH results. However, due to the use of standardized procedures we were able to assess these dissimilarities between the methodology of the project MATCH study and ours, resulting in fair and careful conclusions. In addition, results from the COMBINE study show that for a maximum of 50% of cases “good clinical outcome” was reported at 1-year follow up; 46.8% for behavioral treatment without medication and 50.4% for behavioral treatment with naltrexone (Anton et al., 2006). These figures are not fully comparable with our figures, due to different assessment instruments and follow-up times, but show that even in carefully designed trials, at 1-year follow-up half of the patients are in need of more or extended treatment.

An important question, elaborating on the meaning of the results of CBT in routine practice, is whether these results are satisfying. In considering substance use disorder as a chronic condition, we can compare the results from our alcohol study with studies on other chronic diseases such as depression, hypertension or diabetes. McLellan et al. mention similar “relapse” rates for patients with hypertension and diabetes: “approximately 30% to 50% of patients with Type 1 diabetes and 50% to 70% of adult patients with hypertension or asthma
experience recurrence of symptoms each year to the point where they require additional medical care to re-establish symptom remission” (McLellan, Lewis, O’Brien, & Kleber, 2000).

The contribution of ROM to the learning organization

The final research question of this thesis was whether implementation of a ROM system contributes to the learning capacity of employees. Therefore, we developed an instrument based on the theory of Senge’s Learning Organizations (Chapter 5). In the newly developed Questionnaire for Learning Organizations (QLO), the proposed five-factor structure was revealed using confirmatory factor analysis. However, two of the five QLO scales – Mental Models and Systems Thinking – showed to be in need of improvement.

It was concluded that the QLO measured the proposed five-factor structure, but that some items need to be revised and the scales Mental Models and Systems Thinking need to be extended with one item. Next, efforts should be put into criterion and discriminative validity.

The QLO measures learning capacity on an individual level. However, Senge stresses that teams form the core of learning, and not individuals. It is, therefore, an important question whether the measurement of Team Learning should be performed on individual subjects. To overcome this problem we made the team the object of evaluation in this scale. However, it is still an individual assessment about a group phenomenon.

In the further development of the QLO, comparisons with instruments developed in fields outside the health care sector or based on different theories can be worthwhile. However, the number of instruments with known psychometric properties is low (Moilanen, 2001).

In the study reported on in Chapter 6, we investigated whether learning capacity of professionals increases after aggregated feedback is given about outcomes of the provided treatment, using the QLO. We found no increase in learning capacity after the provision of feedback. This observed absence of learning can be explained by different factors: Psychometric problems of the QLO or the presence of test-retest effects. Another explanation might be that the frequency and intensity of feedback sessions was not enough to create a learning environment for treatment professionals. The conclusion of these two studies is that the ROM system with this frequency and methodology of feedback did not result in an increase in learning capacity, measured with an instrument in development.

Both studies are important contributions to the field of quality improvement and the role of learning organizations in health care. Instruments based on
Senge’s theory are scarce, but recently an instrument became available, measuring learning capacity in teachers in technical and business high schools (Park, 2008). Future studies should be directed towards the relation between learning capacity and quality of delivered care, or even further, treatment outcomes.

Final conclusions

Reviewing the five studies presented in this thesis, it can be concluded that it is feasible to implement ROM with telephonic follow-up interviews. However, the data collected through this procedure suffer from selection bias, which prevents firm valid conclusions on aggregated treatment effectiveness. Moreover, although key persons in the organization appreciated the implemented ROM system, the lack of practical support for the system turned out to be a complicating factor. The ROM database allowed for the generation of conclusions about effectiveness, but the considerations about selection bias, which could result in an over-estimation of positive treatment outcome, should be taken into account. Finally, we found no conclusive evidence about the positive influence of feedback of aggregated treatment outcome data on organizational learning, as measured with the newly developed QLO.

Considerations

All studies in this thesis took place in the setting of substance abuse treatment. The question, therefore, is whether these results can be extrapolated to other settings, like mental health care. Considering the nature of other psychiatric diseases as depression, this is most likely the case. Settings are comparable as well as types of outcome measures. Next, the studies mentioned in this thesis are limited to ROM applied in the aggregated strategy. Another question, therefore, is whether these results are applicable to concurrent ROM. Previously, data collected in concurrent measurements was often used in an aggregated manner for program evaluations (de Beurs & Zitman, 2007; Zwanepol & De Groot, 2008), suggesting that conclusions and suggestions from our study are also applicable for theory, research and practice of that kind of ROM.

Further suggestions for research and practice

The description of the preceding chapters and the chapters themselves contain several suggestions for practice and future research. In the following paragraphs these suggestions are elaborated upon.
Data collection independent from treatment staff

Since data collection is a big burden on treatment professionals, it is vital that this process is organized independently from the treatment process. One of the results of the study on facilitating and impeding factors was to remove the paper and pencil measurement carried out by treatment staff. Collection of data by an independent call center yielded data collected in a feasible manner, although selection bias was a problem. However, this strategy is preferable to strategies where treatment staff are involved, potentially resulting in low response rates, data collection fatigue and lack of reliability that is difficult to assess when treatment staff is involved. The latter issue was also mentioned in a recent study of the harmonization of a monitoring system in Switzerland (Maffli, Schaaf, Jordan, & Guttinger, 2008).

Methods of sample selection for different ROM strategies

In employing ROM for accountability and the creation of an effectiveness database, the selection of a representative sample of patients is a way to overcome the limitations presented for the study in Chapter 2. If concurrent ROM is not the main goal, it is a waste of energy and resources to interview every single patient. However, most SATCs would welcome the implementation of a type of concurrent ROM as well, where outcomes are supportive in individual decision making for the individual treatment process. In doing this, it is recommendable to consider the aggregated and concurrent ROM strategies as separate logistic processes, with possible separate measures to be performed. In concurrent ROM, emphasis can be more on measures that assess processes that can be influenced during the treatment process, like the working relation between patient and clinician and symptom reduction. In the case of data collection for aggregated ROM, measures of interest might be more of the evaluative type, like waiting time, overall satisfaction and quality of life, in addition to symptom reduction. It would be very conceivable to implement ROM with a concurrent “arm”, in which data collection takes place tailored for individual decision-making during and after treatment (i.e. the follow-up interview). The follow-up interview then serves as “Extended Case Monitoring” or “Recovery Management Check-Up” as mentioned before and interviews would be dedicated solely to measures that are relevant for that purpose. Next, an aggregated “arm” can be set up, where a representative sample of patients is interviewed after treatment has ended, thereby collecting information for program evaluation and accountability. In the Netherlands this would imply collecting the outcome data of the mandatory set of performance indicators of the Dutch
Health Care Inspectorate (Inspectie voor de Gezondheidszorg, 2006), next to measures that are of interest for the SATC itself. This design would reduce efforts and costs for the follow-up interview, since all patients would participate in the concurrent ROM arm, where interviews would be tailored exactly to their treatment situation. A representative (random) sub-sample of these patients would also be captured in the aggregated follow-up arm, where the follow-up interview would be extended with questions dedicated to program evaluation.

**Data for concurrent ROM as a by-product of the treatment process**

As we saw in our research project, and as is mentioned in other ROM studies, a valid, reliable real-time outcomes database is still an unachieved goal (Harrison & Asche, 2001; Maffli et al., 2008; Marsden et al., 2008; Teruya et al., 2006; Tiet et al., 2006), mainly because data collection puts big burdens on the treatment staff. Apparently, treatment professionals do not perceive the information they collect as added value for the treatment process. This is an important signal, since concurrent ROM is designed to support decision-making. Whether this has to do with training, the supposed data collection fatigue or disenchantment amongst clinicians remains the question here. The study of Teruya (Teruya et al., 2006) found that staff became frustrated because ROM procedures “[...] took time away from what they valued most – working directly with clients”.

An ideal setting to investigate whether it is possible to collect data for concurrent ROM, without any burden on treatment staff, is Internet treatment. In the Netherlands, several SATCs have implemented this type of treatment, serving a distinct population of patients (Blankers, Kerssemakers, Schramade, & Schippers, 2007). Since all treatment activities are conducted over the Internet, all information regarding the treatment, including outcomes, are digitally entered and processed. Therefore, this information can be regarded as a by-product of the treatment. Editing this information to concurrent or aggregated feedback can support professionals without the extra burden of data collection. So, concurrent ROM of Internet treatment can serve as an important pilot project that could be extended into face-to-face treatment settings.

**Recovery management checkups**

Follow-up interviews can also function as an intervention as was mentioned in several studies (Breslin, Sobell, Sobell, Buchan, & Kwan, 1996; Dennis & Scott, 2007; Scott et al., 2005; Stout et al., 1999). The follow-up interview
can function as a form of aftercare and as such will fit into the stepped care approach (Sobell & Sobell, 2000) where patients in need of a more intensive treatment should be identified as early as possible. This strategy – also referred to as “Recovery Management Checkups” or “Extended Case Monitoring” – has been missed out until now in the redesign process of the Dutch SATC sector (Schippers et al., 2002). In setting up this intervention, it should be considered that the RMC interview could function as data collection for aggregated ROM as well, as was outlined earlier in this section.

Interpretation of feedback of outcomes

As ours and other studies show, lack of knowledge, few feedback mechanisms, inadequate support, resistance, and uncertainty about the purpose of the collected data can hamper participation in the use of outcome data to enhance quality of treatment (Eagar et al., 2005; Harrison & Asche, 2001; Maffli et al., 2008; Marsden et al., 2008; Teruya et al., 2006). Dealing with and interpreting information from outcome data – concurrent and aggregated – should become part of the curriculum of treatment professionals as well as (higher) management. Although the results of our studies on learning capacity are very preliminary, it seems that merely supplying feedback without any training does not result in a learning organization. This suggests that the learning process is not automatic, and therefore should be trained and enhanced.

In addition, training should also be dedicated to data collection. This should be considered a skilled task and should be treated as such, spending considerable time, effort and training acquiring this skill.

Measurement and feedback in different ROM strategies

In setting up a learning curriculum it is important to consider concurrent and aggregated feedback as separate, but parallel processes, just as in data collection. Therefore, the figure with different feedback strategies as introduced in Chapter 1 can be helpful (Figure 1). Measurements in concurrent ROM can be short and frequent. It is suggested that the form (computerized or manual) of the immediate feedback – also referred to as “reminders” – has no influence on the effectiveness of the intervention. However, computerized feedback works best when no pre-emptive action of the treatment professional is required and when a response to the feedback is required (Bywood, Lunnay, & Roche, 2008). In a case where access to feedback was voluntary and where pre-emptive action was required, adherence to this behavior easily decreased after the introduction of feedback systems (Forman et al., 2007).
Measurements dedicated to aggregated ROM can be less frequent in nature – it is even conceivable that a baseline and one follow-up measurement will suffice – and can be fed back after a time interval. In addition, feedback on team level should be provided, but some also advocate aggregated feedback at the individual professional level. In this strategy, each individual treatment professional receives aggregated results of his or her patients (Bywood et al., 2008). Procedures according to the Plan-Do-Check-Act cycle are necessary to improve performance and or outcome. Enriching feedback with benchmark data, as was also the case in our studies, is a strategy that can be effective in improving practice as well (Kiefe et al., 2001).

Figure 1: Feedback strategies in ROM

**Time and effort for data management**

In ROM projects, much is expected from computerized technology, or electronic patient records (EPRs). In our research project, the expectation that EPRs would serve as an easy, accessible source of outcome data did not materialize. Not only did the system fall short as a reliable source of contact details, it also turned out to be problematic retrieving admission information from the system. Historically, the substance abuse treatment sector has been
slow in its adoption of computer technology (Ducharme, Knudsen, & Roman, 2005). This is one explanation why system capabilities fell behind demand. In addition, reliability was not secured, something that was also stated in a report on implementing the earlier mentioned Swiss database, where many different “hands” enter data in the systems (Maffli et al., 2008). This is, however, not unique to the addiction treatment sector. In the sector of cardiac surgery, maintenance or validation of an outcome database was shown to be difficult as well (Fine, Keogh, Cretin, Orlando, & Gould, 2003). Thus far, using EPRs for outcome monitoring is a difficult enterprise where much patience, caution and effort is required.

Promising developments are web-based applications that are specially designed for outcome monitoring. Patients and clinicians can enter information in databases specifically built for this application, providing validity checks and delivering tailored feedback reports. In the Netherlands, NetQ and PsyQuest are the leading systems. Internationally, scientific reports of the application of web-based monitoring systems in different health care settings are starting to become available (Bottle & Aylin, 2008; Forman et al., 2007; Gringras, Santosh, & Baird, 2006; G. T. McMahon et al., 2005; W. M. McMahon, Illmann, & McGinn, 2006; Trivedi et al., 2007). Results are promising, although all the above considerations still hold.

The relation between learning capacity and quality improvement

So far, empirical evidence of the relation between feedback and quality improvement is scarce (Bywood et al., 2008; Forman et al., 2007; Jamtvedt, Young et al., 2006; Thomson O’Brien et al., 2003). Bywood concluded in her recent review on the effectiveness of reminders and feedback in the mental health care and the substance abuse treatment field, that reminders (similar to concurrent ROM in our definition) can have a positive effect on the treatment process, but that the evidence for the positive influence on treatment outcomes was very small or absent. Delivery of aggregated feedback of process variables influenced treatment practice positively, but only when adherence to guidelines for treatment practices was low. Studies on the association between feedback and treatment outcomes were very scarce, and if this was studied no association was found, implying that feedback of process or outcome variables did not result in improved outcomes for patients (Bywood et al., 2008).

Therefore, we proposed in our studies that learning capacity could be a pre-requisite for quality improvement, and that the absence of improved practice or treatment outcomes could be explained by the absence of improved learning capacity. Further development of the QLO and establishing the relation
between learning capacity and quality improvement are valuable steps to be taken in research on outcomes feedback and quality improvement in health care, and more particular in substance abuse treatment.

Research on feedback and learning in substance abuse treatment

Bywood et al. stated that research on the effectiveness of feedback in the field of substance abuse treatment and mental health care was very scarce; only 5% of the reviews in the search were in this field (Bywood et al., 2008). Few studies with satisfying quality were found with regard to the effect of benchmarking in feedback. Moreover, studies that were related to substance abuse treatment were performed in preventive care services, so conclusions are mainly based on evidence collected outside the substance abuse treatment field. In our project, we provided benchmark data, although we did not study the effectiveness of this addition. Results from other studies are promising (Kiefe et al., 2001), but it is not clear whether the same results will be obtained in the field of substance abuse treatment.

In the future, research on other strategies to improve clinical practice should also be studied to see whether they are applicable to feedback of outcomes in addiction care, where the “pay for performance” movement is one of them (Greene & Nash, 2008). This movement represents a strategy where treatment providers are rewarded for pre-established targets in the delivery of health care treatments9. Although there are a lot of reasons to support and to argue against this strategy, it is applied in the United States and the United Kingdom. With the introduction of the mandatory set of performance indicators by the Netherlands Health Care Inspectorate (Inspectie voor de Gezondheidszorg, 2006), the activities of big health insurers, and the reform of the reimbursement system into the “diagnosis-treatment combination” (where a set price for a certain intensity of treatment is reimbursed) the first steps towards a system of pay for performance are being taken.

9 See also: http://en.wikipedia.org/wiki/Pay_for_performance_(healthcare)