Factors for Successful Improvement of Swedish Healthcare

Jesper Olsson

Stockholm 2005

Medical Management Centre,
Department of Learning, Informatics,
Management and Ethics
Karolinska Institutet
SE-171 77 STOCKHOLM
Sweden
Abstract

The Swedish OCM, developed by an Integrative Group Process, was found to be a valid model able to distinguish successful from unsuccessful organizations in terms of improvement. A majority of healthcare organizations applied the Internal Collaborative strategy which lacks the patient centered task alignment characterizing those organizations predicted to be successful by their relatively superior Swedish OCM score. Managers tend to overestimate the prospects of organizational achievement and need support in creating more successful change contexts.

The Swedish healthcare system is struggling with financial limitations, an increasing demand for healthcare and the growing possibility to care for and cure more people than ever before. At the same time studies suggest that between 20-50% of healthcare expenditure is due to costs from poor quality and that about 30-40% of patients do not receive care based on current scientific evidence and best practice. Healthcare organizations undertake various improvement initiatives to cope with these problematic situations but only 20-40% of the improvement efforts aimed at changing practice are successful. From this follows that economical investments as well as time, and human resources are wasted and organizational willingness to embrace further necessary change initiatives diminishes.

The Swedish Organizational Change Manager (Swedish OCM) is Bayesian model intended for use in promoting improvement initiatives and increasing their probability of success. The model was developed using the Integrative Group Process, and frames factors which influence the probability of successfully managing improvement initiatives in Swedish healthcare. The model has also been tested empirically based on a survey of Skellefteå Medical Services. Which validated that the model predicts 80% of the successful, and 75% of the failed improvement initiatives. It can thus be used to diagnose weaknesses in improvement initiatives and to guide corrective actions.

Through the Swedish OCM, more became known about what seems important for improvement strategies in the context of Swedish healthcare. However, in order to aid policymakers a broader picture of improvement work in Sweden was needed. Thus, all clinical department and primary care center managers in Sweden was surveyed with respect to their views on improvement. In the data analysis the following predominating
patterns emerged: improvement methods used were mainly different types of meetings and guidelines, and the main areas of improvement involved administrative routines. Drivers for improvement were mainly intra-organizational, i.e., focused on staff needs. Extra-organizational influences such as patients and comparative data were viewed as less important. Managers in most organizations (80%) perceived this strategy as successful. However, there were little data to support the levels of success perceived. This commonly utilized improvement strategy is denoted the Internal Collaborative Strategy (ICS).

This analysis did not reveal other competing improvement strategies. The Swedish OCM was therefore used to stratify the survey material into two different populations. Units with an exceptionally high Swedish OCM score were postulated to possess the prerequisites for successful improvement (n=24) and comprised one population. The other population contained the remaining units with relatively lower Swedish OCM scores and represent the improvement strategies characteristic of the ICS (n=1038). Comparing these two populations revealed an alternative improvement strategy departing from the ICS. The population with high Swedish OCM score put practical emphasis on patient focus, measuring outcomes, feedback of data, inter-organizational collaboration, communication, information, culture, and development of administration and management. Thus these organizations center their attention on behavioral changes of practice, supported by data. This strategy is denoted patient centered task alignment – PCTA strategy.

Finally, using the same dataset, managers’ perceptions of organizational change and improvement were investigated. This was done based on theories of diffusion of innovation. A population of respondents who perceived their outcomes from improvement initiatives as superior was selected (n=845). The analysis indicated that there are three main underlying factors that characterizing managers’ perceptions of improvement. These are: 1) difficulties in the use of improvement practices and techniques 2) conflict between improvement work and the organization and 3) approaches in use are perceived as effective. In conclusion, there is a need to support change initiatives with practical knowledge about how to use improvement concepts, e.g., local adaptation. There is also a need to find strategies to deal with the conflict between daily work and improvement work. Finally, the data indicates that there might be policy resistance in reorientating current improvement strategies, predominantly the Internal Collaborative Strategy, since managers perceived this as an effective approach to current problems.
Acknowledgements

The journey of creating a thesis is somewhat like Star Trek. The crew of Voyager knows their mission: to search for new life forms. They also know where they are. However, they do not know where they are going and they do not know what they will find, and they certainly do not know all the troubles they will run into during their journey. On the other hand, each episode has a happy ending. Well, if you know Star Trek you know something about the journey of research. I am now one pit stop further down the road. And there is time for some new minutes in the logbook, refueling and making some changes among the crew. As captain of the ship I would now like to take the opportunity to express my deepest gratitude and thanks to all of the competent crew members who has accompanied me so far.

Duncan Neuhauser, for your invisible hand, continuous encouragement and support. David Gustafson, for lots of fun and for introducing me to the IGP. Todd Molfenter, for all our novice discussions nurturing our thoughts on the art of modeling.

Staffan Lindblad, my friend, I am grateful for all your support and all our fun discussions, we talk people into sleep, you know. You have been a great supervisor and I am looking forward to mock-up some of the ideas we have lying around.

Mattias Elg, thank you for all our fun and intriguing discussions, they seem to never end. You have been a great friend from the early beginning of my research. I am looking forward to our future collaborations.

Mats Brommels, thank you for all your valuable comments, but most of all, thank you for giving me the great opportunity to join MMC, it has been very important for me.

Göran Tomson and John Øvretveit, I thank you for your wise and important comments and contributions to my manuscripts.

Anna Bonnevier, for being so helpful, no way I could ever work out the KI administration without you.

Jesper Stenberg, for being a supportive colleague taking care of a great part of our shared work load – I owe you.
Johan Thor, Karin Pukk, Stefan Book, Andreas Hellström and Peter Kammerlind for our extreme research collaboration during the bright nights of spring 2003. It was lots of fun, good discussions and exceptionally much hard work. I really hope we can find an opportunity to do something similar again.

Jonas Lundberg, for the bright nights and for your 24h critical eye services, I appreciate it a lot.

Cheryl Carli for all your instant help reading and commenting manuscripts.

Anders Edström and Ulf Andersson, for all our long discussions and for providing a northern test area.

Bo Bergman, I would probably not have reached this far without your valuable support in the beginning of my research.

Marianne Olsson, Ellen Hyttsten at The Swedish Association of Local Authorities and Regions who believed in me and arranged for the necessary resources for completing this work. I also thank Elisabeth Holmgren for all her valuable support when starting at MMC.

Hans Werthén Foundation and The Sweden-America Foundation for providing the necessary resources to complete the early stages of my research.

My greatest debt is to Axel, Maja and most of all, my wife Ann-Sofi who has given me encouragement and support throughout the project.

Finally, I hope the crew stays on the ship for some further adventures in the galaxy of knowledge…
Original papers

The thesis is based on the following original papers:


Olsson, J., Elg, M., Lindblad, S. “Revealing Specific Characteristics of Change in Health Care Organizations by Stratifying them according to Factors Predicting Successful Improvement” Submitted to Journal of health organization and management.


The papers will be referred to by their roman numerals I-IV

All previously published papers were reproduced with permission from the publisher.
Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Current challenges of healthcare</td>
<td>1</td>
</tr>
<tr>
<td>THEORIES OF CHANGE</td>
<td>3</td>
</tr>
<tr>
<td>Assumptions guiding the present work</td>
<td>6</td>
</tr>
<tr>
<td>Reviews of studies of theories on improvement and change in healthcare</td>
<td>8</td>
</tr>
<tr>
<td>AIMS AND OBJECTIVES</td>
<td>12</td>
</tr>
<tr>
<td>THE IGP METHOD, RESULTS AND COMMENTS</td>
<td>13</td>
</tr>
<tr>
<td>The Integrative Group Process</td>
<td>14</td>
</tr>
<tr>
<td>Phase 1 – Pre-work</td>
<td>14</td>
</tr>
<tr>
<td>Phase 2 – Expert panel meeting</td>
<td>18</td>
</tr>
<tr>
<td>Phase 3 – Empirical tests of the model</td>
<td>23</td>
</tr>
<tr>
<td>Something old something new – the scientific roots of the IGP</td>
<td>28</td>
</tr>
<tr>
<td>Conclusions from developing the Swedish OCM model with the IGP</td>
<td>30</td>
</tr>
<tr>
<td>SUMMARY OF ORIGINAL PAPERS</td>
<td>31</td>
</tr>
<tr>
<td>Paper I - Developing and Testing a Model to Predict Outcomes of Organizational Change</td>
<td>32</td>
</tr>
<tr>
<td>Paper II - Surveying Improvement Activities in Health Care on a National Level – The Swedish Internal Collaborative Strategy and Its Challenges</td>
<td>32</td>
</tr>
<tr>
<td>Paper III - Revealing Specific Characteristics of Change in Health Care Organizations by Stratifying them according to Factors Predicting Successful Improvement</td>
<td>33</td>
</tr>
<tr>
<td>Paper IV - Perceptions of Improvement Work in Swedish Health Care: Implications for Improvement Practices</td>
<td>34</td>
</tr>
</tbody>
</table>
Explanations and some definitions

The term *quality improvement initiative* is defined as an improvement activity within a system in order to improve some quality parameters of the system. What is considered as a quality parameter may vary depending on the needs of the system. Improvement is a deliberate action which does not happen by chance.

Terms like *initiative* and *effort* will be used interchangeably. They are descriptions of improvement work that is time limited. This does not exclude that the improvement initiative may not be a part of a larger, long-term strategic project without a specific time frame.

If nothing else stated the terms *improvement* and *change* will be used interchangeably.

The term *system* refers to “a set on interacting units with relationships among them. The word ‘set’ implies that the units have some common properties. These common properties are essential if the units are to interact or have relationships. The state of each unit is constrained by, conditioned by, or dependent on the state of other units. The units are coupled.” (Miller 1978).
ETE – Estimate Talk Estimate, a technique used to achieve consensus and understand experts underlying heuristics in the IGP

ICS – Internal Collaborative Strategy, a widely spread improvement strategy in Swedish healthcare

IGP – Integrative Group Process. A process to link research and expert knowledge into a tested usable model of a phenomenon.

NPV – negative predictive value

PCTA – Patient Centered Task Alignment, a specific improvement strategy departing from ICS

PPV – Positive Predictive value

Swedish OCM – Swedish Organizational Change Manager. A model for understanding successful improvement initiatives in Swedish healthcare

American OCM – American Organizational Change Manager. A model for understanding successful improvement initiatives in American healthcare


Introduction

The introduction starts with a condensed overview of current problems in healthcare delivery and improvement. Through this I point towards the need for improvement as well as the need to increase the prospects for success of improvement in healthcare. I then continue by introducing four general theories providing perspectives on how change can be viewed. The idea here is to clarify a set of principal views on change to enhance the understanding of what aspects of change intended to cover in this thesis as well as what perspectives on change that will be uncovered. Based on these theoretical considerations I then state my most important assumptions of improvement in healthcare and thus clarify the limitations of the thesis. The chapter then gets more specific and narrows the scope to a number of systematic reviews embracing what previous research points to as important for improving healthcare.

Current challenges of healthcare

New technologies promising better health outcomes, a growing proportion of elderly patients with greater demands on a healthy life, and escalating healthcare costs are often referred to as the greatest challenges to contemporary western healthcare systems. At the same time international studies conclude that there are opportunities to improve health by addressing shortcomings in healthcare quality such as failures to administer appropriate care, provision of inappropriate services, and errors in healthcare delivery (Berwick 1996; Nolan and Schall 1996; Kohn, Corrigan et al. 2000; IoM 2001; OECD 2004).

Instead of attributing the problems of healthcare delivery to scarce resources these studies suggest that there are opportunities to enhance production efficiency and effectiveness. The authors of a recent literature review maintain that between 20-50% of healthcare expenditures are due to costs of poor quality (Sachs 2004; Cederqvist and Hjortendal-Hellman Ds 2005:7) which is more than the poor quality costs of 9-16% of annual turnover seen in other industries (Sörkvist 1998). This is further supported by studies pointing to the large gap between evidence and practice. For example US studies
suggests that about 30-40% of patients do not receive care in accordance with current scientific evidence (Schuster, McGlynn et al. 1998; Grol and Grimshaw 2003).

Consequently, healthcare organizations undertake various improvement initiatives to cope with the situation. Many of these approaches are based on management principles borrowed from American consultants and industry. The underlying essence of these approaches can often be attributed to the Quality movement (Berwick, Godfrey et al. 1990; Gregory, Hackmeyer et al. 1999; Isuoard 1999; Nugent, Kilo et al. 1999).

However, only 20-40% of all improvement efforts in healthcare are successful (Alemi, Safaie et al. 2001; Berwick, James et al. 2003), a majority are failures (Beer, Eisenstat et al. 1990; Hackman and Wageman 1995; Blumenthal and Kilo 1998; Jarlier and Charvet-Protat 2000). Because of contextual nuances in different countries’ healthcare systems it is likely that international knowledge might not be directly applicable to the Swedish healthcare situation (Olsson, Elg et al. 2003). Underestimations of the difficulties in transferring concepts between countries and industries is one plausible cause of failing change initiatives. From this follows that economical investments as well as time, and human resources are wasted and organizational willingness to embrace further necessary change initiatives diminishes. Another problem is that even if change efforts are successful they often remain local and are not easily diffused to the rest of the organization (Øvretveit and Gustafson 2002).
Theories of change

There is an array of different types of changes. Some of them concern structural changes and others concern the process of service delivery. Some focus on the mental models of people within the organization while others focus on their actions or physical settings. On top of this there are numerous theories about how changes within these different areas can be understood, how they start, spread and are sustained. Some of the theories describe change as a natural process, others as different types of stages. This chapter clarifies the general directions and limitations of current research on organizational improvement based on four perspectives of change.

In their seminal article “Explaining Change and Development in Organizations”, Van de Ven and Poole (Ven and Poole 1995) state that the broad variation in theories explaining change based on concepts from different disciplines often encourages compartmentalization of perspectives. They conclude that it would be desirable to try to integrate the sum of all these different theories into one common framework. In order to uncover underlying theoretical patterns, a search of the interdisciplinary literature, based on the words change and development, was conducted. They identified more than one million articles published on the subject in the disciplines of psychology, sociology, education, business, economics, biology, medicine, meteorology and geography. To cope with this prolific literature, they reviewed about 200,000 titles and perused 2000 abstracts, which lead them to carefully read 200 articles. To assess the actuality of the article by van de Ven and Poole a literature search was performed reviewing all 187 articles that had cited it as a reference. In this review no similar effort to develop or refine the overarching framework by van de Ven et al was found. It is thus concluded that this paper probably remains among the state of the art within its field.

In the study by van de Ven and Poole (1995) they propose four basic motors for explaining and driving change in organizations (see Figure 1). They denote these motors: the lifecycle, the dialectic, the evolutionary and the teleological motor. They also propose 16 different combinations of these four basic motors as a way to explain and understand change from different perspectives. The following is a short description of the four basic motors of change.
The lifecycle motor uses the metaphor of organic growth as a way to explain development. The initial stage is the start-up which is followed by growth, harvest and termination. This model is used in a wide variety of research areas. Van de Ven et al (1995) state that the lifecycle motor and the teleological motor is the most commonly used ones for explaining development in organizations. The lifecycle motor follows a single sequence of stages or phases in which characteristics of events are added cumulatively, thus each new step builds on the previous steps. Accordingly, the lifecycle motor views change as an imminent predefined process, like a program or code that is followed and that regulates the change.

The dialectic motor assumes that there are underlying conflicts or contradictory values that compete for domination and control, which reinforces and induces change. Within this motor, stability and change are explained by the power relationship between different agents or entities. As a thesis is challenged by an antithesis the process results in a synthesis. In time this synthesis may become the prevailing thesis as the process continues. By nature the synthesis is a novel construction that departs from both the thesis as well as the antithesis. Through this process the dialectic motor has the possibility of constructing a novel and creative new synthesis, a win-win situation. However this is not always true, sometimes the antithesis mobilizes enough power to overthrow and replace the current thesis. When this happens it is often referred to as a win-lose situation.

Figure 1 Process theories of organizational development and change, (Van de Ven and Poole 1995), reprinted with permission from Academy of Management Review.
The evolutionary motor stems originally from the work by Darwin on the origin of species. The motor has been adapted for organizational purposes and follows the progression of variation, selection and retention. From this perspective the source of change is stochastic variation that creates novel forms of organizations; it just happens (Campbell 1969). The selection of successful variations occurs principally through competition for scarce resources. As well as the lifecycle motor, the evolutionary motor is prescriptive, in the sense that one can specify probabilities of change of a population based on descriptions of their dynamics. This motor has been used in different organizational studies and can be used when the focus is on the process of variation, selection and retention among numerous organizational entities.

The teleological motor proposes that behavior or action (either individual or organizational) is purposeful (see e.g. Simon 1981). It has its basis in the philosophical doctrine that states that purpose or goal is the final cause for guiding movement of an entity (Van de Ven and Poole 1995). Development is viewed as a continuum of goal formulations, implementations, evaluations and modifications of goals. Within the teleological motor there are no predefined or logical steps, nor predefined rules for how change should be made. The center of attention is instead the necessary functions that must be fulfilled, the accomplishments that must be achieved, or the components that must be built or obtained in order to reach desirable goals. These types of prerequisites can be used to assess if an entity is developing (Van de Ven and Poole 1995). Although teleology stresses the purposefulness of change towards a goal, it also acknowledges that in practice there are limits on action and on what is possible to achieve at any given time. Some of these constraints are given by the context of the entity such as laws, contracts and reimbursement systems. If an entity achieves its goals, it does not mean that equilibrium occurs. Maintaining goals achieved within a changing environment requires a continuous change (Prochaska, DiClemente et al. 1992). Also, since organizations are social constructs there are both changes of organizational context as well as changes among different agents within organizations (Sztompka 1991). Goals are therefore reconstructed based on past actions and changing environmental possibilities and constraints, which may create instabilities or gaps that can guide further development. Since goals are made as social constructs, they may take any direction as deemed as appropriate by those formulating and pursuing the goals. Because of these characteristics, theories resting on a teleological motor cannot specify what trajectory development of an organizational entity will follow. Therefore proponents of these theories can at best list a set of possible paths and the rely on norms of decision rationality or action rationality to understand and explain development (Brunsson 1982).

Each of the four basic motors may be differentiated by the two dimensions mode of change and unit of change. The first dimension describes the continuum from prescribed to constructive change. Lifecycle motors and evolutionary motors incorporate a prescribed mode of change. It is typical of these motors that they tend to create change
within an existing framework and thereby produce variations on the prevailing development path. Due to these characteristics these motors seldom create frame-breaking changes.

The teleological motor, as well as the dialectic motor, are constructive in their character. By their basic nature they can break with basic assumptions or frameworks, and may thus produce highly novel features where the outcomes are unpredictable since development is to some degree discontinuous with the past. Based on this, it arises that the teleological and the dialectic motors strive to diverge from the current order. From this perspective they both represent theories of dynamics. The features of intentionality and the ability to change goals at will make teleological theories inherently emergent and creative. As a result the teleological motor has the capability of projecting fundamental and novel changes within the constraints of the system.

**Assumptions guiding the present work**

The following are my current basic assumptions on healthcare improvement. These assumptions have been formed through my research, by others’ research, by discussing with peers and through participation in practical improvement initiatives with healthcare professionals as well as initiatives on a national policy level, working with top managers and politicians. These assumptions can only represent the way I view improvement for the moment, but nevertheless they are my current assumptions on which I build my reasoning in this thesis.

I believe that one of the paramount goals of all health care is to improve health. The foremost guiding concept is that value of care emerges from what happens in the production of healthcare. Production is what happens over time in the daily actions and connections between staff, patients and support systems. It is within these patterns of actions that results from healthcare, good or bad, emerge. Therefore, focus on changing behavior is necessary, as it is a prerequisite of improvement. From this perspective “…change in the individual organizational member’s behavior is at the core of organizational change and, therefore, any successful change will persist over the long term only if, in response to changes in organizational characteristic, members alter their on-the-job behavior in appropriate ways…” (Porras and Robertson, 1993).

Another assumption is that individuals, or groups of individuals, in healthcare organizations are adaptive, purposeful and that they establish goals (spoken or unspoken) in order to reach ends. Thus, healthcare organizations can be viewed as complex, adaptive systems characterized by rapid development and large variation with respect to patients, technology, and knowledge in relation to the degree of freedom to act among professionals. This creates a considerable need for continuous adjustments of actions to
fit current local conditions, leading to changes that seldom can be seen as linear or rational. Instead, the learning from change initiatives unfolds as an interaction between improvements of daily actions and the response of the system. From this perspective “The action is not like a drug; it is more like a health promotion program where the intervention interacts with surrounding conditions, depends on these conditions for effectiveness, and evolves over time in response to the conditions” (Øvretveit and Phik 2004). Improvement work is viewed as a deliberate and planned activity rather than something that just happens (see e.g. Simon 1981). It follows from the above that possibilities of altering daily actions made by practitioners are influenced by both the organizational history (Lewin 1946; Sztompka 1991) as well as the organizational setting as a whole. Seen in this way, the prerequisites of the structures of the system that aid individuals or groups to attain their goals are of importance. For example, different constraints such as the environment, other actors, resources, knowledge, time and support systems (Garud, 2000), are important for improvement work.

I recognize that there are numerous ways of understanding and describing organizational improvement. The introduction touches on some of these. I am also aware of the fact that choosing one way of viewing the subject matter also implies not choosing another and that “a way of seeing is a way of not seeing” (Poogie 1965), quoted by (Van de Ven and Poole 1995)

However, based on the assumptions stated above, I felt most attracted by the teleological motor as the framework for understanding and explaining change in healthcare organizations.

This should not be interpreted to mean that this perspective is better than other perspectives, just simply that this is the perspective I have chosen to use in order to limit the scope of the research. I make this choice since the content of the teleological motor is well aligned with my basic assumptions. I have also observed that there seem to be many other instances in society that acknowledge the teleological motor as one important driver of improvement in health care. Current Swedish examples of such an observation is the work of the national quality registries, the work within InfoVU, Nyckeltalssamverkan, the interest for balanced scorecards, Väntetider i vården, all the work with different breakthrough collaboratives, the work on Ideal Design and different attempts to process orient organizations etc. I argue that the need to put teleology into practice is also continuously displayed in media and reports. Recent examples of this, according to my interpretation, are the scope of Ansvarsutredningen, Ökad tillgänglighet i vården (2005) and Iakttagelser om Landsting (Cederqvist and Hjortendal-Hellman Ds 2005:7). All these examples implicitly build on the teleological motor of change. Consequently, I propose that there is a need to gain a broader understanding of how one can make this motor work better within a Swedish context.
Reviews of studies of theories on improvement and change in healthcare

Depending on the definition of the subject matter there is a good deal of research available. One of the overarching attempts to create research on how to put knowledge into action was made on the behalf of the Cochrane Effective Practice and Organization of Care Review Group. They made an overview of systematic reviews of interventions to promote the implementation of research findings (Bero, Grilli et al. 1998). From this work they state that “Systematic reviews of rigorous studies provide the best evidence of the effectiveness of different strategies to promote the implementation of research findings”. Their conclusions from this review, which are also supported by more recent studies (Grol and Grimshaw 2003), are summarized in three groups depending on to which extent an intervention promotes behavioral change among health professionals.

<table>
<thead>
<tr>
<th>Consistently effective interventions</th>
<th>Educational outreach visits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reminders</td>
</tr>
<tr>
<td></td>
<td>Multifaceted interventions (a combination that includes two or more of the following: audit and feedback, reminders, local consensus process, or marketing)</td>
</tr>
<tr>
<td></td>
<td>Interactive educational meetings (participation of healthcare providers in workshops that include discussion or practice)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interventions of variable effectiveness</th>
<th>Audit and feedback (or any summary of clinical performance)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The use of local opinion leaders</td>
</tr>
<tr>
<td></td>
<td>Local consensus process (inclusion of participating practitioners in discussions to ensure that they agree that the chosen clinical problem is important and the approach to managing the problem is appropriate)</td>
</tr>
<tr>
<td></td>
<td>Patient mediated interventions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interventions that have little or no effect</th>
<th>Educational materials (distribution of recommendations for clinical care, including clinical practice guidelines, audiovisual materials, and electronic publications)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Didactic educational meetings (such as lectures)</td>
</tr>
</tbody>
</table>
The research by Bero was later placed under scrutiny by Solberg (Solberg 2000) as he noticed that “several aspects of this overview of reviews was striking”:

- The focus is almost entirely on behavioral change by individual clinicians.
- There is no direct mention of the role of system change or the practice organization.
- There is no attention to the change process needed for implementation.

Except for an introductory comment about the difficulty of “Disentangling the effects of intervention from the influence of contextual factors” (Bero, Grilli et al. 1998), there were no more attempts to describe these factors nor to describe to what extent the factors were important in understanding successful change.

Based on these reflections, Solberg conducted a new systematic review of reviews trying to identify to what extent these reviews acknowledge the value of system changes, organizational support for change, practice environmental factors and the use of a change process. His main research question was “How do we manage the change process and the organizational factors that control our ability to implement systems?” (Solberg 2000). For this purpose he identified “47 good-quality systematic reviews”. The majority of collected studies have their origin in the U.S., Canada and the U.K. No study has a Nordic or Swedish origin. From this work Solberg concludes that not one of these studies really took a systems approach and that

“There has been little attention to the impact of practice systems or organizational support of clinicians behavior, the process by which change is produced, or the role of the practice environmental context within which change is being attempted. New attention to these issues may help us to better understand the process of improving medical care delivery”

From these systematic reviews, Grol and Grimshaw (2003) as well as Solberg (2000) conclude that there is a need for more research to gain better insight into the important processes and elements of successful change. Which strategies work? They also acknowledge that it is possible to change behaviors, but that this change generally requires comprehensive approaches at different system levels i.e. doctor, team, practice, hospital and wider environment. Solberg’s conclusions are aligned with the theories of teleology, which define the constraining factors of the surrounding systems as important factors in explaining the formulations of targets in organizations as well as organizational capabilities to change and achieve targets.

These systematic reviews lead to the knowledge area of organizational change and similar areas such as diffusion of innovations. As mentioned in this chapter’s introduction, there are lots of theories available for understanding different aspects of change. One of the recent, large systematic literature reviews on the topic was done by Greenhalgh et al (Greenhalgh, Glenn et al. 2004). They scanned over 6000 abstracts and identified over 1200 full text papers and over 100 books and book chapters that were
possibly relevant. From this material some 495 were included in the study, of which just 213 were based on empirical analysis of data, whereas the remaining majority were non-empirical.

Through this systematic review they identified 13 research areas that, largely independently of one another, had provided evidence relevant to the diffusion of innovations in health service organizations. In Figure 2 they give a limited overview of their findings. Each heading in the figure points to a research area contributing to the rich body of knowledge on the topic of change. The aim of the figure is to point to the large number of diverse perspectives that could be used to understand and explain change and diffusion of innovations.

![Figure 2 Conceptual Model for Considering the Determinants of Diffusion of innovation, Dissemination, and Implementation of Innovations from (Greenhalgh, Glenn et al. 2004) reprinted with permission from the author.](image)

Despite the in-depth review conducted by Greenhalgh et al, they state that they purposely excluded one other relevant area in the organization and management literature - organizational psychology. Here there are relevant theories in which innovativeness is seen as dependent on good leadership, sound decision-making, and effective human resource management (especially the motivation, training, and support
of staff). Thus their model describing perspectives of research on change would actually be even larger if including these perspectives too.

It is obvious that the emerging picture of the number of possible perspectives on change and diffusion of innovations is extremely broad, as concluded by van de Ven et al (1995). Some of this research is empirically based but most of it is probably not. Some of this research is conducted within healthcare settings; most of it is probably not. A small fragment of this research might originate from Swedish healthcare, but the vast majority does not. Hence there is a need to compile the variety of theories into something useful, such as a model of change. It is also desirable to empirically test such a model. Finally it is desirable to contribute to the understanding of improvement within a Swedish context.
The aim of the thesis is to support policy makers and researchers in the field by exploring, based on insights from theory and empirical observations, conditions for success in improving Swedish healthcare.

The specific objectives are:

- to develop and empirically test a decision model that predicts and explains successful improvement initiatives in Sweden
- to explore the accuracy and usefulness of the Integrative Group Process to compile research knowledge and experience
- to generate an overview of and to analyse the improvement strategies used in Swedish healthcare
- to analyse improvement strategies used in Swedish healthcare based on a teleological framework
The IGP method, results and comments

Since one of the specific objectives of this work is to gain insights on the practical usefulness of the Integrative Group Process (IGP) as a method for building decision theoretic models, it is appropriate to have a somewhat deeper description of the method than is possible in a peer review journal. The section will therefore start with a selected history of the IGP to be continued with a step by step description of the process. The section ends by clarifying some of the links between IGP and other research techniques.

Previous research similar to this research has been undertaken by, among others, Professor David H Gustafson and Dr Todd Molfenter, University of Wisconsin, Madison. They developed and tested an American Organizational Change Manager (American OCM) (Gustafson, Cats-Baril et al. 1992; Gustafson, Sainfort et al. 2002; Molfenter, Gustafson et al. 2005). This model was explicitly developed using a research process called the Integrative Group Process (IGP) to explain improvement in American healthcare. The same approach is used in paper I. The IGP method has also been used by the National Health Services (NHS) in collaboration with Professor Gustafson to develop a comprehensive model for explaining sustainability of change in British healthcare (work in progress). This model shows similarities with the American OCM, but has a different aim. Common to both of these approaches is that they concern how to develop comprehensive and empirically tested models to understand complex phenomena. However, both these models are developed to explain change in other countries and they can not a priori be assumed to work within a Swedish context (Olsson, Elg et al. 2003). This research has been undertaken in close collaboration with professor Gustafson, University of Wisconsin, Madison.
The Integrative Group Process

The Integrative Group Process (IGP) used for developing the Swedish OCM model (Figure 3) is a variant of the original process as described by Gustafson, Fryback and Rose 1981 (Gustafson, Fryback et al. 1981; Gustafson 1991; Gustafson, Cats-Baril et al. 1992). This research process may be used to model any Bayesian or value model based on experts’ knowledge. The purpose of the process is to obtain knowledge from a group of experts in such a way that the model developed can predict reality as accurately as possible.

The overall IGP model development process has three main phases as described below. This process generates a model that is ready for use and continuous refinement.

**Phase 1 – Pre-work**

This phase starts with choosing a phenomenon to model and studying the earlier knowledge about the phenomenon, and what kind of model to develop and to start identifying the expert panel.

**Choosing a statistical model**

There are numerous options for how to statistically model a phenomenon. The type of statistics chosen has implications for which data to gather in the IGP, as well as how and in what order data is captured. In the current case a Bayesian odds model approach is used which is the same type of statistics used in the original IGP when developing the
American OCM model. Elaborations on this particular choice are made under the section *The choice of statistics.*

Bayes’ theorem is often referred to as the formally optimal mathematical method for revising prior opinion in the light of new evidence (Laplace 1951; Winterfeldt and Edwards 1986; Gustafson, Cats-Baril et al. 1992). In the current IGP the odds form of the theorem is used. It has three key components:

\[
Posterior\ odds = \text{likelihood ratio} \times prior\ odds
\]

Each term in the equation has an explanation. The starting point is the *prior odds* (also called a priori odds). The prior odds represent the odds of a certain event, before knowing anything about its revisers (see below). For example an odds, in the context of improvement initiatives, would be the probability calculated by dividing the probability of being successful with the probability of an unsuccessful change project.

The *Likelihood ratios*, also referred to as revisers, represents the diagnostic power of new evidence about a certain event. Examples of revisers in the context of improvement might be if it became known that everyone preferred a specific improvement initiative or if resources were allocated for conducting the initiative. Such new evidence would revise the prior odds for success. Adding evidence from all known revisers will adjust the prior odds and results in the *posterior odds* which is the final estimate of the probability for having a successful improvement initiative.

With in the Swedish IGP, the odds form of the Bayesian theorem was deployed:

\[
\frac{P(H_0|f_1, \ldots f_n)}{P(H_1|f_1, \ldots f_n)} = \frac{P(f_1, \ldots f_n|H_0)}{P(f_1, \ldots f_n|H_1)} \times \frac{P(H_0)}{P(H_1)}
\]
### Table 1: Explanations of the Bayes theorem based on the Swedish OCM

<table>
<thead>
<tr>
<th>Symbol:</th>
<th>Meaning:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>Designates the experts’ subjective probability of an event.</td>
</tr>
<tr>
<td>$H_0$</td>
<td>The mutual exclusive hypothesis that a change effort will be successful.</td>
</tr>
<tr>
<td>$H_1$</td>
<td>The hypothesis that a change project will fail.</td>
</tr>
<tr>
<td>$f_{n_i}$</td>
<td>Identifies a change project’s $n^{th}$ factor $f$ to the $i^{th}$ level. The Swedish OCM have 11 factors with 3 levels attached to each factor.</td>
</tr>
<tr>
<td>$\mid$</td>
<td>Designates ’given’; thus $H_0 \mid f$ says ”successful change project given the project has characteristic $f$.”</td>
</tr>
<tr>
<td>$P(H_0 \mid f_{i_1} \ldots f_{n_i})$</td>
<td>Measures the belief that a change project will be successful given the characteristics $f_{i_1}$ through $f_{n_i}$. This variable shows the probabilities of the event to be predicted after examining revisers $f_{i_1}$ through $f_{n_i}$.</td>
</tr>
<tr>
<td>$P(H_1 \mid f_{i_1} \ldots f_{n_i})$</td>
<td>Measures the belief that a change project will be a failure given the revisers $f_{i_1}$ through $f_{n_i}$.</td>
</tr>
<tr>
<td>$P(f_{i_1} \ldots f_{n_i} \mid H_0)$</td>
<td>Is the probability of finding characteristics $f_{i_1}$ through $f_{n_i}$ among successful change initiatives. This variable is called the likelihood associated with the characteristic and is a figure, which sets the impact of the revisers on the forecast.</td>
</tr>
<tr>
<td>$P(f_{i_1} \ldots f_{n_i} \mid H_1)$</td>
<td>Is the probability of finding characteristics $f_{i_1}$ through $f_{n_i}$ among failed change initiatives.</td>
</tr>
<tr>
<td>$P(H_0)$</td>
<td>Is the probability of a change project being successful if one is totally ignorant of whether the project has the revisers $f_{i_1}$ through $f_{n_i}$ or not. This is the probability of the event before any other information about it is available.</td>
</tr>
<tr>
<td>$P(H_1)$</td>
<td>Is the probability of a change project being a failure when knowing nothing about the specific characteristics $f_{i_1}$ through $f_{n_i}$. Note: $p(H_0) + p(H_1) = 1$</td>
</tr>
</tbody>
</table>
Establishing an expert panel

The choice of experts is important to the outcome of the IGP process. The expert panel should represent the area of interest from different perspectives, in this case the knowledge of organizational change from diverse perspectives (Stumpf 1978). The choice of experts sets the knowledge base that is used in the IGP to avoid expert bias. This is done in several ways, although the basic concept is to use different views of a phenomenon to make a deeper exploration of a topic among the expert group. For this purpose it is typically good for the model to include experts with both practical as well as theoretical background in the knowledge field modeled (VandeVen 1974; Gustafson, Cats-Baril et al. 1992). Approximately 5-9 experts are needed. If the group is too small, one person’s view can dominate the discussion. On the other hand, if the group is too large some of the members might not actively participate in the discussions. The key issue is to constitute and use the expert panel in such a way that it elicits subject exploration and avoids polarization since this may lead to bias and an oversimplified model of reality (VandeVen 1974; Gustafson 1991; Gustafson, Cats-Baril et al. 1992).

Since the IGP approach relies on expert knowledge, the choice of experts is important to the outcome. In the Swedish OCM model seven experts representing change management, leadership and quality improvement techniques from both a practical as well as a theoretical standpoint were included. The team of experts comprised three professors: two of them with a special interest in Total Quality Management and improvement, and one with a special interest in human behavior; three change agents from health care: one was a physician and CEO for a governmental body working at the R&D department with organizational improvement issues. The second one was a former R.N. now working both at a strategic and at a practical level with improvement, facilitating improvement teams for a municipality, including healthcare. The third expert had a clinical background from the field of laboratory medicine, and has also been involved in union work and has worked strategically and practically with systems improvement at a midsize hospital over the past 8 years. The final expert was from outside the healthcare industry, a senior CEO with 25 years’ experience of change in large industrial organizations, who also had knowledge about healthcare from working as a consultant.

Interviews

After appointing the experts, the next phase in IGP is to interview them. The interviews are used to construct the straw model, which is the first draft of the model. The straw model is later used in the second phase of the IGP, the expert panel face-to-face meeting. The interviews are both a way to collect data for the straw model, but also to have experts start reflecting on the phenomena being modeled.
Interviews help to diversify content of the straw model, which is beneficial since the objective at this stage of the development process is to explore the knowledge domain (Gustafson, Cats-Baril et al. 1992). There are three general questions the interviewer is interested in: first, the definition of an outcome objective, in the case of the Swedish OCM the definition of a successful change project; second, what factors are of interest to understand if a change project will be successful. This is done by asking the experts: “Suppose you had to predict whether a particular change project was going to be successful according to the stated definition. What would you want to know before making the prediction?” Third, with what indicators could one measure the extent to which a factor is present within a change project or not? This can be captured through a question like: “What questions could give you answers that would make you optimistic about the project? And what would make you pessimistic?” The important part here is “What questions…” these answers (in the format of questions/statements) are used later as indicators in the model. In the Swedish OCM work, interviews rendered 13 hours of recordings transcribed into 70 pages of text. From this material a non-duplicated list of factors and indicators was created and categorized. Factors found by literature review, not identified in interviews, were added to this list. This rendered a straw model containing 22 factors with some 3 to 9 indicators in each factor.

Phase 2 – Expert panel meeting

Expert panel meeting

The expert panel meeting is a two-day face-to-face workshop. The broad process for capturing data at this meeting is; 1) to decide on the objectives of the model i.e. the definition of successful change, 2) based on the straw model, identify factors and indicators most important for predicting/understanding a successful change initiative, 3) estimate the prior odds for the objectives and 4) determine the model factors and indicators associated with likelihood ratios needed to create accurate posterior odds, and finally, 5) test the model against hypothetical profiles. This agenda aims at obtaining data for developing a model according to Bayes theorem (see page 15).

The two-day face-to-face experts’ meeting began with introducing the straw model
(Figure 4) a picture of the room layout when briefing the experts). The first objective was to agree on a definition for the desired outcome (i.e. what the model was to forecast), which for the Swedish OCM model became:

“At the end of the project, the changes made have positive and prolonged effects regarding social/human aspects and/or objectively measured improvements. The new solution is a part of the daily work and meets future demands.”

Thereafter, the expert panel started reviewing the straw model by increasing the number of factors using the Nominal Group Technique (a method for generating ideas and creating affinities using, for example, post-it notes) followed by decreasing and merging factors until there was a non-redundant model agreed upon.

The next phase in the experts meeting is to decide on indicators for each factor, (Figure 5). The indicators are signs used to understand to what extent a factor is present in a change initiative or not. They are constructed as statements or questions, which you either can agree to or not. Here they are written as affirmative sentences (Table 2). There are in general two types of factors. Those who have an explanatory value and those that is actionable. Experts decided to include both types in the current model. In the Swedish OCM model, experts decided to use four indicators for each factor, representing three levels of impact on the a prior probability.

Table 2 an example of a factor from the Swedish OCM model. ‘Tension for Change’ is the factor name. The four indicators below measure the presence of this factor in a change project. For example, users check boxes that four indicators are met. This would equal the high level of performance, which according to Table 3 translates into the odds 4.92 to be used in the Bayesian calculation.

<table>
<thead>
<tr>
<th>Tension for Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff are dissatisfied with the way things are being done because they do not work.</td>
</tr>
<tr>
<td>Customers are dissatisfied with the current system.</td>
</tr>
<tr>
<td>We are being out-performed by others.</td>
</tr>
<tr>
<td>Budgets are continually being cut and fewer people are being asked to do more with less.</td>
</tr>
</tbody>
</table>

These levels are: high (4 indicators met), medium (2-3 indicators met) and low (0-1 indicators met).
Experts now had to address the model with their subjective probabilities to develop the prior odds and the likelihood ratios. To obtain the prior odds for success, the experts answered the following question on a form:

“Based on your experience, what proportion of change projects turn out to be successful according to our definition?”

A typical question to obtain a subjective probability for the numerator portion of the likelihood ratio in the Swedish OCM model would be, regarding for instance the three levels of the factor Tension for Change:

“Suppose you had a pile of 100 randomly selected successful change projects. How many of these would have the factor ‘tension for change’ set at a low level? Medium level? High level?”

To obtain the denominator portion of the likelihood ratio, the question posed is:

“Suppose you had a pile of 100 randomly selected unsuccessful change projects. How many of these would have the factor ‘tension for change’ set at a low level? Medium level? High level?”

The latter two questions were repeated to create likelihood ratios for the three levels of all eleven identified factors. All experts’ likelihood estimates were collected on special data sheets after individual assessments. After the individual assessment, a group discussion takes place based on the Estimate-Talk-Estimate technique. Here data from datasheets were displayed on a white board and estimates with a wide spread were discussed among the group (Figure 6). After each discussion, experts independently have an opportunity to re-estimate the power of the levels. This process helps identify bias and allows experts to change their statements if needed. It is also a tool to enhance researchers’ understanding of how the experts are thinking about the model.
During the described integrative group process the experts assessed the different parts of the model with subjective probability measures as a means of providing input for discussions about their heuristics of the model.

Until this phase of the IGP experts have generated the prior odds as well as the subjective estimates for all included factors three levels (low, middle, high) of impact to the prior odds (Table 3). Thus, a first testable expert generated model has been developed.

**Table 3 displays the sum of success and failure for each factor’s three levels. These data are then divided into sum success/sum failure to generate the mean subjective likelihood ratio for each of the three levels of each factor.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>sum success</th>
<th>sum failure</th>
<th>mean odds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shared platform</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>295</td>
<td>60</td>
<td>4.92</td>
</tr>
<tr>
<td>Middle</td>
<td>225</td>
<td>235</td>
<td>0.96</td>
</tr>
<tr>
<td>Low</td>
<td>80</td>
<td>305</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>Problem analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>280</td>
<td>40</td>
<td>7.00</td>
</tr>
<tr>
<td>Middle</td>
<td>230</td>
<td>160</td>
<td>1.44</td>
</tr>
<tr>
<td>Low</td>
<td>90</td>
<td>400</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Improvement context</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>320</td>
<td>46</td>
<td>6.96</td>
</tr>
<tr>
<td>Middle</td>
<td>190</td>
<td>160</td>
<td>1.19</td>
</tr>
<tr>
<td>Low</td>
<td>90</td>
<td>394</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Tension for change</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>295</td>
<td>60</td>
<td>4.92</td>
</tr>
<tr>
<td>Middle</td>
<td>235</td>
<td>175</td>
<td>1.34</td>
</tr>
<tr>
<td>Low</td>
<td>70</td>
<td>355</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>Leadership for improvement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>350</td>
<td>130</td>
<td>2.69</td>
</tr>
<tr>
<td>Middle</td>
<td>185</td>
<td>150</td>
<td>1.23</td>
</tr>
<tr>
<td>Low</td>
<td>65</td>
<td>320</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>Colleagues participation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>275</td>
<td>110</td>
<td>2.50</td>
</tr>
<tr>
<td>Middle</td>
<td>235</td>
<td>225</td>
<td>1.04</td>
</tr>
<tr>
<td>Low</td>
<td>80</td>
<td>265</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>305</td>
<td>70</td>
<td>4.36</td>
</tr>
<tr>
<td>Middle</td>
<td>225</td>
<td>165</td>
<td>1.36</td>
</tr>
<tr>
<td>Low</td>
<td>70</td>
<td>365</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>The project team</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>315</td>
<td>105</td>
<td>3.00</td>
</tr>
<tr>
<td>Middle</td>
<td>195</td>
<td>205</td>
<td>0.95</td>
</tr>
<tr>
<td>Low</td>
<td>100</td>
<td>300</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>Learning and reflection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>290</td>
<td>115</td>
<td>2.52</td>
</tr>
<tr>
<td>Middle</td>
<td>225</td>
<td>225</td>
<td>1.00</td>
</tr>
<tr>
<td>Low</td>
<td>85</td>
<td>260</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>Resources</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>305</td>
<td>55</td>
<td>5.55</td>
</tr>
<tr>
<td>Middle</td>
<td>200</td>
<td>155</td>
<td>1.29</td>
</tr>
<tr>
<td>Low</td>
<td>95</td>
<td>390</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Project launch</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>275</td>
<td>80</td>
<td>3.44</td>
</tr>
<tr>
<td>Middle</td>
<td>235</td>
<td>165</td>
<td>1.42</td>
</tr>
<tr>
<td>Low</td>
<td>90</td>
<td>355</td>
<td>0.25</td>
</tr>
</tbody>
</table>
First tests of validity and reliability

The expert panel’s next step was to generate data for the first tests of the inner reliability of the model. The test consists of the experts predicting the probability of success for a set of hypothetical profiles. These probabilities are later compared to the posterior odds generated by the model. This is a fast and cost-effective, and if this test fails, it is unlikely that the model could predict reality. In the current process, each expert assessed 72 hypothetical profiles (Figure 7).

Figure 7 The table displays the first 21 of the total of 72 hypothetical profiles generated for the first test of the model. The numbers (0 and 1) in the matrix represent the presence (1) or not (0) of a certain indicator within a specific factor of a hypothetical profile. The example at the bottom of table illustrates what the 16th hypothetical profile’s first 12 indicators looked like. All 72 profiles in the table were translated via a database into a similar format as the example at the bottom of table. The profiles were put in a random order and assessed by the expert panel through the question: What are the chances of success for a project looking like this? (0 to 100%). The individual experts’ answers for the first 10 profiles were discussed through ETE as a way to calibrate an expert’s judgment, before the remaining 62 profiles were assessed.
There are many ways to produce profiles. The most used is the random approach, although when developing the Swedish OCM, an \(I_{12}\) Design of Experiment (DOE) approach was used.

This first test of the model aims at investigate how experts’ assessments of the profiles correlate against the model’s assessment of the profiles. This test only informs if the model is able to predict the experts (who have developed the model). If the model cannot pass this test it needs to be remodeled. On the other hand, if the model seems to be good enough at predicting the experts, one may consider this as a confirmation of internal validity and continue with further empirical tests.

A linear correlation between the Swedish OCM model and the hypothetical profiles gives \(R^2=0.73\). Thus the model fit equals about 88%. This result was interpreted as fair enough, confirming that the model is ready for further empirical tests.

**Phase 3 – Empirical tests of the model**

In the original IGP process as described by Gustafson and Fryback, two main options of validating the model developed are suggested: 1) test the model with a second group of experts, 2) conduct an empirical prospective or retrospective study.

In this particular case it is suitable to continue testing the Swedish OCM model in two other ways: 1) small-scale tests with two different breakthrough teams focusing on the perception and usefulness of the model, and 2) a larger survey of Skellefteå Medical Services focusing the model’s ability to discriminate between successful and unsuccessful change initiatives.

**The empirical test of the Swedish OCM**

The aim of the first empirical tests was to improve the linguistics of the model towards a format that is easy to understand, but still reflects the intentions of the experts.

The first tests of the model were conducted in two different settings with teams participating in a breakthrough collaborative. The first collaborative that tested the model was working with the subject ‘waits and delays’. The aim of the test was to get a better understanding whether practitioners interpreted the indicators in a similar way to that intended by the experts when formulating the Swedish OCM model. The initial test was done by giving teams (one team contained about 3-5 people from the same organization) representing 8 different organizations the Swedish OCM to use the model as a survey at one of their meetings. The specific task was to write down when they felt uncertain about what was meant with a question. Based on the comments in this material the linguistics of the model were revised.
The second empirical test of the model was conducted at a primary care center that was working with a breakthrough collaborative on waits and delays. In this homogenous test (all respondents from the same small organization) about 10 different teams of 3 people each assessed the model in two steps: first individually and then as a group. The group wrote down their comments to the questions in the model and handed in the data. The researcher was also available during the session for discussions. Based on the comments from the groups and the discussions with different people, the linguistics of the model were revised a second time according to the original Swedish OCM model generated by the experts.

So far the IGP generated a model containing 11 factors important for successful improvement. Following is a summary of the model (full model available in paper II):

**Resources**: there are sufficient resources for the project; time, knowledge, money

**Problem analysis**: There are data illuminating the importance and the potential in solving the problem. This makes sense for those involved.

**The project team**: there are channels for dialogue between the improvement team, staff and management team. The project team have mandate to change necessary parts of the system undergoing improvement.

**Project launch**: there is a clear and communicated project plan and the management team has been involved in developing aims and stating that improvement is desirable.

**Feedback**: feedback is given for learning purposes. There are outcome variables that rapidly can indicate if a change is an improvement, or not, this is communicated by respected persons.

**Learning & reflection**: there are possibilities for local adoptions of aims and ideas and for training of staff. There is time for reflecting and learning, there is also joy in the improvement work.

**Improvement context**: persons who think differently are seen as assets and critical persons are met with respect. Risk taking and experimenting for learning purposes is encouraged.

**Colleagues' participation**: information about the improvement work is easy to access and readily available. There are arenas, used by everyone, to discuss and influence the improvement initiative.

**Shared platform**: there is a shared view about how to work with improvement and what the aim and vision are.

**Leadership for improvement**: leadership have the ability to prioritize. They ask how things are going and are not afraid of letting others lead the improvement initiative.

**Tension for change**: there is a clear improvement potential through dissatisfaction over the performance and/or outcomes of the system. There is a visible potential since other comparable organizations have much better outcomes.
The third empirical test focused to test model’s positive predicted value (PPV) and negative predictive value (NPV). For this purpose a large survey containing 260 questions was distributed (N=1460) to all of the staff in Skellefteå Medical Services (survey available in appendix 1). The large number of questions in the survey was due to a joint research effort to collect data both on the Swedish OCM model as well as on leadership styles. One of the ideas was to test if any particular leader characteristics can be associated with a climate in favor for successful improvement. However, this is not the scope of the current topic.

The response rate on the survey (n=638) is considered fair with respect to the in-depth questionnaire. To enhance the understanding of nonresponders a question in the survey was compared with a similar question in a staff survey used in Skellefteå Medical Services. This comparison indicated that respondents were representative of the entire staff population of Skellefteå Medical Services. In all, 138 respondents were excluded since they had 12 or more unanswered questions in the survey. This leaves us with a dataset for analysis containing n=500 respondents. These data represent all units of Skellefteå Medical Services with exception from six small units.

As described earlier the Swedish OCM model contains a set of questions which respondents are to answer yes or no to. Because the scope of the Skellefteå Medical Systems survey was broader than only testing the Swedish OCM model, this forced the use of 7 grades Likert scales. Since such scales have a middle point they are difficult to transform into being dichotomous. For example, if a respondent answers ‘4’ would that rate as a yes or a no on the original Swedish OCM dichotomous scale? This was resolved by adding the score of the answers of each indicator within a factor. This summed score has then been divided into three equally distant categories. For example if the factor contains 4 indicators, the maximum score possible is 7*4=28. This score has then been divided by three (as many levels as the experts have estimated for each factor). This gives the equidistant of 9.33 between each of the three subjective odds levels in the factor. In this case the three levels low, middle and high would be translated into the scores 1 to 10 =low, 11 to 19 =middle and 20 to 28 =high. By this procedure each respondents answer to all included factors has been translated into the three factor levels and then each specific level has been substituted with the subjective odds for the present interval of the specific factor (see for example table 3). This later procedure is the same as when using the dichotomous scale. The score from each respondent has then been compared against respondents' perceived success of finished improvement initiatives undertaken the last 12 months.

Typically for odds models is that they tend to generate a wide spread, non-linear function. The Swedish OCM model generates an odds ranging from 0 representing the lowest possible odds to 4978511 representing the highest possible odds. In these situations it is customary to compress data. There are different methods for this
procedure. In the Swedish OCM model one of the regular compressing techniques, log odds transformation, was used (Gustafson, Cats-Baril et al. 1992; Clemen 1996).

This data was used to test PPV and NPV for the Swedish OCM model. Since such test is a binomial function there is a need to decide cutoff points where it is possible to say if an improvement initiative was a success or failure and when the model should predict an improvement initiative as successful or a failure. This is an arbitrary process where the levels are chosen with respect to what kinds of errors that are tolerable. For example it is possible to set the cutoff limit for the model so high that it only captures successful improvement initiatives. However, such a cutoff would on the other hand make the model to falsely identify an increased number of actual successes as failures. How one chooses to optimize the cutoff limits is linked to the intended use of the model. In this case the model is optimized to identify projects in need for help in order to become successful. This then renders a greater risk helping improvement initiatives that probably would be successful anyway. I thought it was worthwhile to help too many compared to the high price of a failed improvement initiatives.

Accordingly the cutoff for success or failure was set to 1.7 (Figure 8). Such a limit gives a model that predicts 80% (PPV) of the successful improvement initiatives correctly and falsely identifies 20% as successful improvement initiatives when they are not. Further, it identifies 75% (NPV) of the unsuccessful improvement initiatives correctly but falsely identifies 25% of the unsuccessful improvement initiatives as potential successful ones.

![Figure 8 Respondents log odds compared to perceived success. The horizontal line indicates the Bayesian model cut-off value.](image-url)
Correlating Swedish OCM against staff and patient surveys

Every fourth month Skellefteå Medical Services conducted a staff survey and a patient survey among all their departments. This had been done just before the Swedish OCM survey was distributed. This fact gave us the opportunity to investigate to what extent Swedish OCM data correlates with patient satisfaction and staff satisfaction. This should not be regarded as a test of the model, rather as testing for linkage between climate for successful improvement and staff- and patient satisfaction. Since the unit of analysis in staff and patient surveys was the organizational unit rather than single respondents as in the described survey, there is a need to aggregate the Swedish OCM measures to the level of an organizational unit. This was done through calculation of the arithmetic Swedish OCM mean score for each unit. This arithmetic mean Swedish OCM score was then correlated with the mean score for each units answer to the two other surveys. Through this procedure a strong positive and significant correlation between a Swedish OCM score and patient and staff satisfaction was identified. This may be interpreted as when the Swedish OCM score increases (a climate in favor for successful improvement) staff and patient satisfaction increases (more satisfied patients and staff).

<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th>p</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish OCM and employee satisfaction</td>
<td>0.41</td>
<td>0.02</td>
<td>26</td>
</tr>
<tr>
<td>Swedish OCM and patient satisfaction (in-patient units)</td>
<td>0.73</td>
<td>0.02</td>
<td>6</td>
</tr>
<tr>
<td>Swedish OCM and patient satisfaction (out-patient units)</td>
<td>0.72</td>
<td>0.02</td>
<td>8</td>
</tr>
</tbody>
</table>
Something old something new – the scientific roots of the IGP

Based on the description of the IGP and the findings about the usefulness of the Swedish OCM model it is of some interest to broaden the understanding what scientific methods that could be traced within the IGP. First there is a need to acknowledge that the aim of the IGP method is to effectively develop a model that is able to predict something of the phenomenon modeled. And, since the method builds on use by humans, it needs to be designed to handle different human biases efficiently enough to produce accurate models of real phenomena. Some of the customary biases the modeling effort has to handle are the following:

<table>
<thead>
<tr>
<th>Availability</th>
<th>The easier it is to remember a certain event, the higher the risk of overestimating the frequency of the event (Tversky and Kahneman 1973). Hence, recent events tend to be overestimated (Spetzler and Holstein 1975).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selective perception</td>
<td>People are predisposed to see information that is aligned with their views of life (Dearborn and Simon 1967; Spetzler and Holstein 1975).</td>
</tr>
<tr>
<td>Frequency bias</td>
<td>People tend to have strong recollection of impressive events and assume they occur more frequently than they actually do. At the same time, they underestimate the frequency of common, casual events (Einhorn and Hogarth 1979).</td>
</tr>
<tr>
<td>Anchoring and adjustment</td>
<td>People make predictions by starting with an opinion about the event most likely to occur (anchoring) and then hesitate to move away from this opinion as new information becomes available (adjustment). (Kahneman and Tversky 1972; Kahneman, Slovic et al. 1982).</td>
</tr>
<tr>
<td>Illusory correlation</td>
<td>The belief that two variables co-vary when they do not (Einhorn and Hogarth 1979).</td>
</tr>
</tbody>
</table>

Once aware of these potential biases, there are different techniques built into the IGP to minimize the impact of them on a model. These fundamental components of the IGP can be traced to several traditional research approaches, such as:

*The Delphi technique* which is a non-face-to-face procedure for aggregating group members’ opinions (Dalkey and Helmer 1967; VandeVen 1974; Gustafson, Cats-Baril et al. 1992). In this technique, members answer questionnaires, and the researcher compiles the answers and sends them back to respondents for comments. A modified version of this is used at the beginning of IGP when generating the straw model, except that
telephone interviews are used instead of questionnaires, and data is fed back via the straw model. Parts of the Delphi technique are also used in the end of the IGP when assessing subjective probabilities to the final model. This is utilized through the Estimate-Talk-Estimate approach (described below). This technique reduce customary bias by providing experts an opportunity to reconsider their answers in the light of others knowledge contribution.

Nominal group techniques. This is the generic name for face-to-face group techniques that consist of steps like: silent idea generation, sharing of ideas, clarifying and discussing ideas, individual reassessment, and ranking of ideas (VandeVen 1974; Gustafson, Cats-Baril et al. 1992). Parts of this approach are used in IGP when compiling the straw model into the final model. It is also used when experts assess priorities of ideas combined with discussion and re-estimations. This is utilized in the Estimate-Talk-Estimate (ETE) technique, which is a method to discuss model inaccuracies that may be caused by expert bias. ETE improves accuracy by forcing group members to consider the views and experiences of others and then independently revise their opinion based on what they have learned (VandeVen 1974; Gustafson, Cats-Baril et al. 1992). This diminishes customary bias errors.

Social judgment theory is a group process based on cognitive feedback (Hammond and Rohrbaugh 1976), i.e. trying to understand the heuristics behind different answers. The procedure of the method is to ask subjects to rate a series of scenarios. This is used to provide statistical feedback identifying the importance of different components in the scenarios. An initial consensus model is built and used to predict a new series of scenarios, noting instances in which the model is inaccurate. Within IGP this is used throughout the process, but also after the experts’ meeting when conducting initial test of the model’s ability to predict the experts. This reduces customary bias by providing researchers an opportunity to look for variations in expert assessments of scenarios. When inconsistency or large differences appear it provides researchers with an opportunity to get back to experts for clarifications on their judgements.

Group communication strategy originates from a set of normative instructions (Hall and Watson 1970) stating that conflict reducing techniques such as voting and bargaining should be avoided along with behaviors like arguing and win-lose statements if the objective is to develop high-quality judgments where the group will later accept the outcome (Nemiroff, Pasmore et al. 1976). The IGP facilitator uses this knowledge to maintain a productive dialogue among group members during the expert panel meeting.

As described previously, there are several potential sources of expert bias. The usefulness of the IGP process as a method for successfully overcoming expert bias has been proven in several different settings. Examples of this include: quality of care modeling (Gustafson and Sainfort 1990), suicide predictions (Gustafson, Greist et al. 1977; Geisler and Rubenstein 1987), index of medically underserved (Gustafson 1975),
the American Organizational Change Manager (Gustafson, Cats-Baril et al. 1992; Gustafson, Sainfort et al. 2002) and most recently medication adherence and sustainability of change. The IGP research model has also passed validity and reliability testing (Gustafson, Edwards et al. 1969; Gustafson, Fryback et al. 1983). The IGP is considered effective in avoiding bias and provides researchers with detailed and elaborated input for developing different value models (Gustafson, Cats-Baril et al. 1992).

**Conclusions from developing the Swedish OCM model with the IGP**

Through the described IGP process a decision-making model has been developed. The model was developed by combining research with expert opinion.

Empirical tests conclude that the model predicted 80% of successful and 75% of unsuccessful improvement initiatives in Skellefteå Medical Services. It has also been shown that high model scores indicating a climate in favor for successful improvement initiatives, correlates significantly with higher staff satisfaction and with higher patient satisfaction. Thus the model may be used to diagnose weaknesses in improvement efforts, to measure an organization’s overall potential for successful outcome of improvement and to prioritize potential initiatives under consideration. Finally the study confirmed the usefulness of the Integrative Group Process for producing accurate predictive models in a complex environment.
Summary of original papers

Following is a summary of original paper I through IV. The chapter starts with an overview in.

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Population</th>
<th>Contribution</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.</td>
<td>Survey of total population</td>
<td>All clinical department managers (n=958) and primary care center managers (n=1355), Survey of entire population (no selection)</td>
<td>A framework describing prevailing general improvement strategies in Swedish healthcare</td>
<td>January to June 2003</td>
</tr>
<tr>
<td>III.</td>
<td>Survey of total population</td>
<td>Same as study II. Data stratified into two data sets based on Swedish OCM score. Data sets compared against each other</td>
<td>Clarifying the difference between the prevailing general improvement strategy and an empirically identified strategy closer to teleology</td>
<td>January to June 2003</td>
</tr>
<tr>
<td>IV.</td>
<td>Survey of total population</td>
<td>Same as study II. Selection of respondents who perceive their improvement work as very successful</td>
<td>Clarifying some of the problems of the prevailing general improvement strategy based on a management theory</td>
<td>January to June 2003</td>
</tr>
</tbody>
</table>
Paper I - Developing and Testing a Model to Predict Outcomes of Organizational Change

The aim was to develop and empirically test a model to predict chances of conducting successful improvement, and to explore the accuracy and usefulness of the Integrative Group Process.

For this purpose a Bayesian prediction model was built through IGP. Relevant research on improvement and a group of theoretical and practice experts were used to develop a first model and perform initial tests of internal validity. The models positive and negative predictive value were then empirically tested on data from a survey of all staff in primary care, emergency, hospital, specialty and ambulatory care at Skellefteå Medical Services. Response rate was 41% which rendered n=638 respondents. From these respondents 138 was excluded since they had twelve or more unanswered questions in the questionnaire. This leaves a material containing n=500 respondents. Comparison of nonresponders against controls indicated that responders were representative for the entire staff population of Skellefteå Medical Systems.

The results and conclusions from the study are that the Swedish Organizational Change Manager (Swedish OCM) developed predicted 80% of successful, and 75% of unsuccessful improvement initiatives in Skellefteå Medical Services. It has also been shown that high model scores i.e. a climate in favor for successful improvement initiatives, correlate significantly with higher staff satisfaction and with higher patient satisfaction. Thus the model can be used to diagnose weaknesses in improvement efforts, to measure an organization’s overall potential for successful improvement and to prioritize potential initiatives under consideration. Secondly, the study confirmed the usefulness of the Integrative Group Process for producing accurate predictive models. More detailed description of the IGP is found in chapter The Integrative Group Process.

Paper II - Surveying Improvement Activities in Health Care on a National Level – The Swedish Internal Collaborative Strategy and Its Challenges

Through the Swedish OCM (paper I), more became known about what seems important for improvement strategies in the context of Swedish healthcare. However, in order to aid policymakers a broader picture of improvement work was needed. Some of the specific questions regard the fields in which improvement activities are undertaken and by what technologies? What are the drivers of, and barriers for, change? And to what extent are improvement initiatives perceived as successful?
For this purpose data from a survey including managers of all primary health care centers (N = 958) and all clinical hospital departments (N = 1355) in Sweden was analyzed through descriptive statistics and total response rate was 45.9% (n=1062) (survey available in appendix 2). A qualitative analysis of non-respondents indicated that respondents seem to be more interested in the topic of the survey and that non-respondents mainly come from small units.

The results and conclusions from this study indicate that the predominant improvement technologies used were different types of meetings and guidelines. Main areas of improvement were administrative routines. Drivers for improvement were mainly intra-organizational i.e. focused on staff needs. Extra-organizational factors such as patients and comparative data were viewed as less important. Most organizations (80%) perceived this strategy as successful. However, there were few organizations having data to support the levels of success perceived. This commonly used improvement strategy is denoted the Internal Collaborative Strategy (ICS).

**Paper III - Revealing Specific Characteristics of Change in Health Care Organizations by Stratifying them according to Factors Predicting Successful Improvement**

In the previous paper we concluded that the ICS is the prevailing strategy of change. However, this study only looked at general patterns of change. If one acknowledges that there probably are competing strategies in use, these patterns might be identified by stratification with a set of critical success factors as operationalized through Swedish OCM.

In order to conduct such an analysis data from the ICS study was stratified into two separate populations based on respondents’ Swedish OCM scores. Based on this procedure one population representing Swedish OCM top scores was identified and the remaining respondents represented the general Internal Collaborative Strategy perspective. These two populations were then compared against each other with respect to significant differences in how they have answered the questions in the survey. Swedish OCM questions used for stratification were excluded in the comparison.

Results indicate that the Swedish OCM top population differed from the comparison population at the p.05 level in a number of the surveyed characteristics. Major differences were that the Swedish OCM top organizations put emphasis on patient focus, measuring outcomes, feedback of data, inter-organizational collaboration, learning and knowledge, communication/information, culture, and development of administration and
management. Thus these organizations center their attention on behavioral changes supported by data. This strategy is denoted *patient centered task alignment – PCTA strategy.*

**Paper IV - Perceptions of Improvement Work in Swedish Health Care: Implications for Improvement Practices**

From the dataset used in study II n=845 respondents claiming outstanding outcomes from improvement were selected. Their perceptions of nine statements based on theories of diffusion of innovations were analyzed by correlations to perception of results and through a factor analysis (including Varimax rotation) intended to identify any underlying factors of perception.

The analysis indicated that even though the study population claimed that ‘the combined results of all your improvement work have been worth the effort’ there were only vague (but significant at p.01 level) correlations between four of the nine statements and the perceived outcomes from improvement. It was also found that there were weak connections between the perception of improvement and the view whether improvement yield ‘more healthcare per monitary unit’. One of the significant correlations was between ‘small scale tests’ and the perception of ‘more healthcare per monitary unit’.

The factor analysis identified three underlying factors explaining perceptions of improvement. 1) managers believe that the current approach is an ‘effective approach that meets perceived needs’ 2) Half the population claims there are ‘difficulties in the use of improvement practices and techniques’. 3) Approximately half the population state that there are ‘conflicts between improvement work and the organization’.

Finally, the data indicates that there is a need to create support and guidance in improvement work and that there might be policy resistance in reorientating current improvement strategies, predominantly the Internal Collaborative Strategy, since managers perceived this as an effective approach to current problems.
Discussion

The work in this thesis has operationalized a teleological perspective on healthcare change theory by developing a Bayesian model for predicting the success of improvement initiatives. Knowledge of previous research and practical experience with change in healthcare has been integrated into a comprehensive model called the Swedish Organizational Change Manager. The Swedish OCM has been validated by empirical tests in a Swedish context and shown to have a positive predictive value of 80%.

The chosen method of developing models of complex phenomena, the Integrative Group Process, is well established internationally. Here, its application has produced a prediction model valid and specific within the Swedish context.

The current improvement work in Swedish healthcare has been surveyed and a predominant strategy of internal collaboration satisfactory to leaders has been observed but the strategy is seldom oriented towards either the needs of patients or towards external partners or stakeholders. An alternative strategy was detected at healthcare units predicted by the Bayesian model to be successful. This strategy was clearly patient centered and aligned with the tasks of healthcare. Furthermore, the strategy was a good example of the comprehensive improvement methods recommended in extensive literature reviews of evidence based methods for improvements.

This work combines management theory and the practical experience of experts with empirical data to produce new knowledge that might be applicable in future improvement efforts in Swedish healthcare. However, there are methodological considerations to be made in the discussion below as well as identifying what this thesis contributes to existing research before suggesting how practitioners and leaders of healthcare in Sweden may apply this knowledge contribution to solve their context specific practical problems.
Methods used

This section concerns some of the methodological considerations within this thesis, starting with the Integrative Group Process and ending with considerations of the survey methodology.

Why replicate the American study?

An alternative to replicating the whole IGP research done by Gustafson et al would have been to translate and test the American OCM model within a Swedish context. The main argument for not doing the latter is researcher's interest in testing and learning how to use the IGP process. Another important consideration is the researcher's close collaboration with Professor Gustafson, one of the developers of the IGP method, which provided unique possibilities to learn with the source. Also, conducting a new IGP under Swedish conditions has probably provided a richer understanding of factors influencing improvement in Sweden, which in turn provides a large bank of ideas for further research.

The choice of statistics

The Swedish OCM model could have been based on a different statistical assumption. The choice of statistical method has an impact on how to adapt and use IGP, as well as on how to improve and use the model in practice. Some alternative statistical methods therefore need to be addressed along with arguments for why the Bayesian statistics were the preferred choice.

Common alternative methods used in the IGP approach are to gather data for developing 1) multi attribute utility (MAU) models 2) Bayesian Belief Nets (BBN) 3) structural equations models and 4) conjoint analysis models.

MAU models are linear and additive. They tend to be good predictors when all factors influencing an outcome are known, which many times is difficult to fulfill (Gustafson, Cats-Baril et al. 1992). Bayesian models on the other hand are multiplicative and their strength resides in the ability to revise probability estimates based on new information. A disadvantage is that small errors in estimates may magnify to significant errors in the model, since they are multiplicative. However, the benefits of multiplicative models include low sensitivity to missing factors, i.e., it is possible to develop good predictive models without knowing everything about the modeled phenomena. However, the major argument for a Bayesian approach lies in the fact that this is a replication study. This is reinforced by historically successful applications of Bayes’ theorem in similar situations (Greist, Gustafson et al. 1973; Gustafson 1975; Gustafson, Cats-Baril et al. 1992). These
things guided the choice of Bayesian statistics for this study. However recent research based on the IGP method have been utilizing structural equations models (SEM) and conjoint analysis. SEM is a way to analyze data and to set factors of a model in relation to each other, testing the power of the assumptions of the model constructed. This approach can probably be used on any IGP developed model. Since I argue for the need to understand change from a more holistic perspective, it would be intriguing to develop a SEM model based on current empirical data. When considering the use of conjoint analysis, preliminary results suggest that this method actually has the potential to produce event better models than the approach used in current research. This method was however recently developed when NHS built a sustainability model but the approach was not available when constructing the Swedish OCM model.

**Usefulness of experts**

There are different ways to capture data for the construction of statistical models describing reality. A proven useful method in traditional health care modeling efforts is to rely on records or clinical databases. However, these approaches have some disadvantages when developing value models in general. Medical records tend to be a poor source of knowledge since they often lack the information needed to develop a model (Gustafson, Edwards et al. 1969; Greist, Gustafson et al. 1973). Also, this approach is often time consuming and costly since many medical records have to be reviewed in order to generate enough statistical power in the sample to be able to develop a model. Databases, on the other hand, constrain opportunities for developing novel models since the type of data collected are already determined when constructing the database. This can lead to the development of flawed models missing important predictors or misrepresenting selected predictors (Greist, Gustafson et al. 1973). Also, the use of databases and records relies on the idea of high power predictable probability frequencies (stable or trends) identified through many cases. Then, a substantial shift in mean value (or other tendency measures used) based on a newly introduced better technology, will take many new cases to become detectable (Gustafson, Fryback et al. 1983). This may cause the use of registries and clinical records to be insufficient data providers in a fast evolving environment such as health care.

As a complement to the traditional data sources, research demonstrates that experts’ knowledge can be used to model different phenomena and aspects of healthcare as well as other complex situations. Research also indicates that this approach can be a more cost and time effective way than traditional model developing methods (Gustafson, Edwards et al. 1969; Greist, Gustafson et al. 1973; Gustafson, Fryback et al. 1983; Gustafson, Cats-Baril et al. 1992).
From the research it may be concluded that with the right techniques, expert generated models are as consistent with current reality as traditional modeling styles (Greist, Gustafson et al. 1973; Gustafson, Fryback et al. 1983; Spiegelhalter, Myles et al. 1999). It is argued that the subjective models’ strong predictive performance resides in their ability to model the complex and irrational decision heuristics present in ordinary human decision-making (Winterfeldt and Edwards 1986; Kahneman et al 1982) as well as modeling the specific contextual nuances of a given decision or prediction (Lilford and Braunholtz 1996). The strengths of using experts also indicate that that IGP models based on experts might quite possibly have a specific niche. It may be particularly useful in relation to especially complex situations where rational as well as irrational factors are important for making a model useful. With respect to how to gather data for constructing a future model, combining experts’ opinions and traditional modeling styles may prove to be a successful approach. Such an approach could open up opportunities to model complex areas, which has not been practicable before.

The surveys

There are several of ways to gather knowledge about patterns of improvement in healthcare. However the single most popular method used within organizational studies comprises questionnaires (Fulop, Allen et al. 2001; Czaja and Blair 2005). They range from being very formal using only Likert scales to being less structured, relying on open ended questions. Surveys are well suited for descriptive studies, but can also be used to explore aspects of a situation, or to seek explanations and provide data for testing hypothesis (Kelly, Clark et al. 2003) (Fulop, Allen et al. 2001).

Some of the characteristics and advantages of surveys, according to Kelly et al (Kelly, Clark et al. 2003) and Czaja et al (Czaja and Blair 2005) are that they:

- Produce data based on real world observations (empirical data)
- Mail surveys has been proven successful in the collection of data about sensitive topics
- The breadth of coverage of many people or events means that it is more likely than some other approaches to procure data based on a representative sample, and can therefore be generalizable to a population
- Surveys can produce large amounts of data in a short time at a fairly low cost. Researchers can therefore set a finite time span for a project, which can assist in planning and delivering end results
- Surveys typically only produces a ‘snapshot’ of how things are at a specific point of time
Based on these kinds of issues the survey alternative was chosen in this research. It was also believed that none of the other available methodological options would, in a time and cost effective manner, be able to produce a reliable nationwide improvement pattern (paper II through IV). This choice was also guided by a short timeframe and economical constraints, all in combination with the funding organization’s wishes. As in most research, different constraints make perfection unrealistic which brings us into some practical considerations with respect to some of the known disadvantages with surveys.

To be able to understand the impact of different sources of error to the results of a survey study there needs to be some kind of definition of what a good survey outcome might be. According to (Czaja and Blair 2005) a good survey is one that; “is a valid measure of the factors of interest, motivates respondents to cooperate and elicits acceptably accurate information on the topic of interest”. From this somewhat broader definition, the tradition in survey literature has been to focus on the accuracy of the survey results as a measure of a good survey. Accuracy is often expressed in terms such as the frame, non-coverage, sampling, non-responders, measurement error, and processing error (Kalton 2001; Platek and Särndal 2001).

The frame

The frame is the scope of the population that is able to answer the questions in the survey. In this case the discussion is about which respondents could answer the questions posed in the survey? It is feasible to believe that both staff, managers, CEOs, politicians and patients would, from some perspective, be able to answer the questions posed. However, it was thought that CEOs and politicians would not be able to describe actual patterns of improvement practice in healthcare. Another interesting group would then, of course, be staff. However, this group was not thought of as practically doable frame as it would create extensive problems with sampling strategies as well as getting the questionnaires to the respondents, i.e., having access to an up-to-date address registry. It was then decided that middle managers would be a suitable level for representing the organizational level of interest. It was also possible to get hold of these people with a reasonable amount of work.

The choice of middle managers also deserves some further elaboration. It is not believed in general that middle managers will be able to answer question truly objectively. Hence, it can only be expected that managers would produce answers based on their interpretations of reality and the context in which they work. Taking this argument to its very endpoint, it is only possible to capture their perceptions of reality. To some degree these perceptions will align with an objective reality and to some degree it won’t. The assumption was, that despite a high or low degree of alignment between reality and perceived reality managers would act according to their perceptions of reality – since this is what they think is real. Thus it is of interest to know what their perceptions are.
In this papers II through IV there are some problems with the frame since the address database included units who should no longer be in the database. This problem leads us to overestimate the total number of units eligible for the survey, which makes the response rate falsely low.

**Non-coverage**

Non-coverage is where potential respondents within a given frame can not be reached, for example when no address or telephone numbers can be found. The problems of non-coverage generally receive little attention, probably because it is almost impossible to measure. Yet non-coverage can be a greater concern than non-response, particularly for some population subgroups (Kalton 2001). In the current studies there is probably no great problem with non-coverage even though some departments probably were missed.

**Sampling**

There are numerous bodies of literature on different sampling techniques and their advantages as well as their disadvantages. These issues are extensively covered in most textbooks on surveys. The main argument for using sampling is that it is less expensive than a survey including the total population (Czaja and Blair 2005) and is warranted when surveys of total populations is unrealistically expensive. Sampling is the technique of taking a smaller sample thought to be representative for the whole population of interest and apply statistics to the sample in order to be able to infer something about the total population. From this perspective, the use of sampling statistics introduces uncertainties and bias, as compared with studying an entire population. Since study II – IV study the entire population framed and thus does not use sampling, the art and problems of sampling is not an issue here. However, it can be concluded that any problems that are related to survey accuracy in a study of the total population are probably more complex and worse in a sampled study.

**Non-responders**

Non-responders are respondents who do not answer the survey. It is easy to produce non-response rates but they can only serve as indicators of the potential for bias (Kalton 2001). Non-responses potentially contribute to bias in survey estimates if their answers differed from responders. Of course, typically it is not known if respondents differ from non-respondents on a survey measure. This is because by definition, there are no data from non-responders. On the other hand, it is almost always useful to try to determine the reasons for non-response. If telephone numbers are available, a call to a sample of the non-respondents to determine their reasons for not returning questionnaires may be helpful (Czaja and Blair 2005). In the current studies this method has been used to enhance the understanding of non-response. The survey manager conducted 50
telephone interviews with non-responders. These interviews reveal that most non-responders represent smaller units. The reason given for not responding was often lack of time.

There is also additional knowledge about non-responders in general. For example, typical non-responders have low education, do not like to write and/or read, and/or have no interest in the topic (Czaja and Blair 2005). Based on this, it is possible that the main reason in these studies for non-response is lack of interest in the topic. It is not too remote to link the issue of low interest in a subject matter to the degree of attention paid to the subject matter in practice. Consequently, it is quite possible that the ones who have answered the survey are respondents with a greater interest in the topic of improvement, which in turn raises speculative ideas of what non-responders might have answered. Another contribution to non-response is if the survey is too difficult or time-consuming. The questionnaire we used was large, time consuming and, compared to many other questionnaires, difficult. This has probably contributed to some of the non-response. On the other hand, this reinforces the fact that respondents really are interested in the subject and thus prioritize answering.

Processing error

Processing errors are those errors that could occur at the various stages of processing data for analysis. Such errors can arise during data entry, editing, coding and calculations. The potential contribution of such errors to bias is seldom calculated, and has not been done in current work. However, for the tasks of data entry, editing and coding, a professional company working with the issues daily was used. The use of experts may be viewed as a precaution to avoid processing bias.

The potential bias introduced in calculus was dealt with as a team effort. The studies in this thesis along with some additional studies based on the same data material were used within a collaborative of researchers. Within the collaborative, researchers independently used the same data. This gave us the opportunity to continuously check the consistency of the results against each other. This procedure probably avoids some of the potential calculus errors.

The attention paid to these various sources of error in designing surveys and evaluating their estimates is generally geared more to what can be easily measured and reported, than what is important (Kalton 2001). Thus sampling errors and non-response receive a good deal of attention, were as other sources of error are often largely ignored (Kalton 2001). Some of the more fundamental problems are, for example, poorly formulated questions, inappropriate response scales, exclusion of key variables etc (Thompson 2003). How respondents interpret questions scales, real life situations, as well as the context of the respondent when answering the questionnaire, the fact that respondents only answer what is socially accepted, memory issues, and so forth are all potential
sources of error. In itself, just the process of taking a measurement is a subject to error. Such measurement errors are, of course, not restricted to surveys but are found throughout the empirical sciences. In essence many of these problems are about choices that must be made between for example measuring a desired concept directly but inaccurately or measuring a related concept accurately. Some of these potential sources for bias can, to a certain extent, be controlled or understood, others not. They are all issues of relevance when trying to judge the accuracy of surveys.

All these sources of errors are often a major cause of inaccuracy in survey estimates, but they are hard to measure and hence are rarely quantified. Special studies are needed to understand the validity and reliability of survey responses. Since such studies generally require the investment of substantial resources, they are more likely to be carried out in relation to repeated surveys, when the study may lead to improvements in future rounds of the survey, rather than in connection with ad hoc surveys (Kalton 2001).

The longstanding tradition in the survey literature has been to focus on the accuracy of the survey results, where accuracy is generally expressed in terms of low sampling and non-sampling errors. The concept of total survey error has been advanced, with the idea that surveys should be designed to minimize the total error in the survey estimates by achieving the optimum balance between various error sources (Kalton 2001). One thing to do in order to cope with these issues is to try to write questionnaires that will convey to respondents a uniformly understood request for information. Especially mail surveys requires a questionnaire that is totally self-explanatory; the importance of clear and simple statements can not be overstated, because instructions and questions must be uniformly understood by a wide variety of respondents. This task is a difficult one for which there are fewer scientific principles to follow than, for example, for sample design (Czaja and Blair 2005). It seems like there is no easy or straight forward answer to the question if a survey is accurate enough to answer the questions of interest to the investigator. Each survey needs to be scrutinized in the light of its own uniqueness, based on a variety of data. Based on these issues it is reasonable to believe that a high quality survey development process could serve as one type of quality assurance. But even with respect to this experts state that “there can be no universal recommendations on best practice in respect of questionnaire design and survey conduct” (Thompson 2003). Despite this Czaja and Blair (Czaja and Blair 2005) have tried to define what they view as a high quality survey design process (Figure 9). When comparing the design process of the current questionnaire to this process (paper II), it can be concluded that there is a close fit. This is true with the exception for the top process named ‘draft preliminary sampling plan’ whereas sampling was not an issue in study II – IV.
Even though the topic investigated in this thesis is quite complicated as many ambiguous words as possible were kept out of the survey. In cases where jargon was needed the possibility to answer ‘do not know’ or similar was provided for respondents. Much effort was also put into the layout of the survey. We tried to make the instrument as self-explanatory as possible while not making it too compact. This process was supported by practitioners and managers from real healthcare settings, both continuously in the survey development process as well as occasionally, when performing tests of the survey within different groups of managers.

Even though it is not possible to account for all traps in the art of making and conducting surveys the process used is at least coherent with the ideas of how to conduct surveys properly. Within such processes the problems of using surveys are revealed and discussed before deciding on appropriate actions.

**Contributions and further research on the topic**

This research aims to make a contribution to both practitioners and researchers working in Swedish healthcare. One of the contributions to research is to demonstrate the utility of the IGP method also in a Swedish context. The IGP is a cost effective method for developing and testing models of complex phenomena, as it can combine rational as well as irrational aspects of the phenomena. The IGP seems to be widely applicable when there is a need to translate compiled research and experience into comprehensive concepts or actionable models. This feature of the IGP has implications also for
practitioners when trying to put research into practice. From their perspective it is important to find ways to combine local knowledge with for example decontextualized research. Including the right experts, i.e., local opinion leaders, in the modeling process will also help practitioners disseminate a model. In fact, versions of the IGP have already been used in one of the Swedish county councils when they linked their own cases of good healthcare access to research knowledge on the topic. Through this process they generated concepts to use within a local improvement collaborative. This county evaluated their use of the IGP process as successful enough to start a discussion on how it could become a part of their county-wide strategy. Another initiative on the same theme was undertaken by the Swedish Association of Local Authorities and Regions (SALAR), developing concepts to improve Swedish cancer care and concepts to improve care of patients with dementia. The concepts developed were successfully deployed in two national improvement initiatives and SALAR plans for future use of the IGP in similar situations.

When it comes to the further development of the IGP the most intriguing development path could be to combine IGP with the ideas of systems dynamics modeling.

Considering the Swedish OCM model, my intention is that it should be used as a framework for discussions among improvement teams about their abilities to successfully take on an improvement initiative. Through such discussions, they could clarify what they need to do in order to increase their chances of success.

To my knowledge the model has been used this way within a few projects in primary healthcare and within one national measurement project. The feedback from their use was positive. However, I believe that the potential to use the model in these ways can be further developed, for example, by adding action resources to the problematic picture which was uncovered through the use of the Swedish OCM model.

With regards to future studies on the validity of the Swedish OCM there is an on-going project with Uppsala Clinical Research Center (UCR). The aim of this work is to conduct a prospective test based on clinical data.

Finally, the paper describing the Swedish Internal Collaborative Strategy (paper II) has been used in a variety of different national and local conferences. It is estimated that more than 3000 people ranging from top management and politicians in Swedish healthcare to professional practitioners have had the opportunity to listen to and reflect upon the basic nature of the ICS. In this way, paper II has made a contribution to the general Swedish discussion on change. It is worth noting that almost all persons I have presented the material to find it intriguing. They also confirm that it is confirms their ‘gut feelings’ about the prevailing general improvement strategy in Sweden.
Pursuing teleology?

In this final section I take the discussion on the findings one step further. The starting point for this is my argument that teleology is an important piece in the remedy of healthcare. As mentioned in the introduction the basics of the teleological motor is about goals as social constructs and testing changes in the system in order to learn how to reach ends more effectively. This also includes the continuous reformulation of goals. This motor of change is postulated as being constructive and has the possibility to create novel development pathways. My argument is that it seems like many instances of society believe in the teleological motor. Examples of this are that many counties and departments work with different measurement feedback models such as balanced scorecards. Other examples are the discussion of the use of the Swedish national quality registries. In general, the public discussions tend to move toward increasing demands for transparency of data and the formulating of goals to understand the quality, effectiveness and efficiency of healthcare. All these examples utilize the basically teleology idea of aims and measurement to aid improvements of everyday practice.

Based on the presented research what are the obstacles for the spread of the teleological motor of change and how is it possible to increase the returns of using it?

Managers do their best

Deming said “Every manager supposes that he is doing his best. He is, and this is the problem.” (Deming 1986 p. 36). What does he mean? The following is an attempt to elucidate the meaning of the quote from the perspective of teleology.

The importance of this discussion lies in the assumption that management and leadership are crucial components for successful improvements. This assumption is supported by most research and management literature (Deming 1993; Senge, Roth et al. 1999). Then, from the observation that ICS is the prevailing improvement strategy, it follows (see paper II and IV) that managers are quite satisfied with what they are doing and the results thereof. The basics for this assumption is that within the ICS strategy, staff is satisfied, improvement meets norms and values, improvement does not create conflicts and managers are satisfied with their outcomes from improvement projects. This is the result from a strategy based on staff needs, focusing on improving administrative routines by writing documents and talking to each other.

In contrast, the units working with the PCTA strategy, which is anticipated to represent the actions of teleology, were found to put emphasis on patient focus, measure outcomes, use feedback of data and seek extra-organizational collaboration. At the same time they do not work on these issues at the expense of the staff focus. So, if managers
actually start to measure and use data, to set goals and to support actual change in on the
job behaviors they will end up in a delicate position as compared with today.

The problem is why would managers do anything else than what they already are doing?
If they start working on the PCTA strategy they will experience problems that they do
not have today. For example increased transparency with make it evident how well (or
badly) they are doing in comparison to others. Actual change in on the job behavior, i.e.,
doing more than writing papers and talking to each other, will probably lead to more
conflicts among staff than today. Why would managers want that? Also, perhaps many of
today’s management tasks would be substituted with other, new tasks.

Other obstacles for pursuing a PCTA strategy are the potential managerial knowledge
problem linked to for example new managerial tasks. This problem exists because today’s
managers are recruited by yesterday’s managers. Yesterdays managers were experts in
handling and surviving in an ICS influenced system, consequently so are most of the
current managers. How to act as a manager within a PCTA influenced system defines the
knowledge gap. This returns us to the quote from Deming “Every manager supposes
that he is doing his best. He is, and this is the problem.” Many of today’s managers
continue to manage according to the ICS strategy, and they do it well and this is a
problem for the promotion of the PCTA strategy. Hence there is a need to develop
knowledge on what it takes to be a manager in a PCTA strategy, but also what the
specific demands are on organizational structures in order to be able to support the
PCTA strategy.

Finally, I want to call upon the definition of a system as “a set on interacting units with
relationships among them. The word ‘set’ implies that the units have some common
properties. These common properties are essential if the units are to interact or have
relationships. The state of each unit is constrained by, conditioned by, or dependent on
the state of other units.”

If attention is shifted towards other hierarchical levels of healthcare it is necessary to
acknowledge that the middle managers who have answered our survey are part of a larger
system. They too are “constrained by, conditioned by, or dependent on the state of other
units”. Thus the actions of and the signals from top management and politicians need to
be taken into account when considering improvement initiatives. Although it seems as if
top managers and politicians implicitly support the teleological approach, they do not act
in this way and as a result it is possible that their actions might preserve the current
system. For example, they may say that aims and measurement are important but then
they rarely ask how organizations is doing, or only ask about costs. Another example is
using data for making decisions without taking into account that one part of teleology is
the social construct of common aims and directions. Giddens (1976) describes the
problem of all managers as being bounded. "Men produce society, but they do so as
historically located actors, and not under conditions of their own choosing" i.e.
individuals are creative and have a free will but at the same time our history bounds and restricts the alternatives for action. This might be one explanation of the power of status quo, but it also suggests that all levels of the system should collaboratively pursue the teleological approach in order to be successful.

Directions for the future?

Based on the empirical findings in paper I through IV, and considering the teleological motor of change as outlined in the PCTA strategy, the following principles are argued to be of importance for the transformation towards a PCTA strategy.

Principle 1: Organize support for patient centered task aligned improvements

Even among units perceiving themselves as achieving excellent outcomes from their improvement work a vast majority report that it is difficult to apply improvement technologies and techniques (paper IV). This fact focus on the need for the creation of resources intended to aid and guide managers and practitioners in how to translate general concepts into context specific applications that help alter behavior. Such resources should provide profound insights and skills with respect to management and change, both in theory as well as in practice. Managers need to arrange for this support to be readily available if their units are to be successful in patient centered task aligned improvement work.

Conduct change initiatives from the patients’ perspective.

In the internal collaborative strategy staff needs are the most important driver for change (paper II). However, the importance of a patient orientated system is supported in paper III where organizations adhering to the Patient Centered Task Alignment strategy have a significantly higher degree of patient focus. This idea is also further supported by the observation that the PCTA strategy includes an extra-organizational component, trying to meet the demand of patients by paying attention to other parts of the organization. This highlights the fact that patients perceive healthcare as a whole process of care with themselves at the epicenter rather than pieces of care defined by care givers organizational units.

Organize support for professionals working together with patients

In paper III we show that one of the core ideas of PCTA is the focus on task alignment and changes in on the job behavior. Consequently, managers and policymakers need to get interested in trying to understand the prerequisites and the driving forces of improvement in the daily work in order to be able to manage a task alignment strategy. It
is, however, important not to confuse paying attention to, and understanding daily work, with detailed governance of activities of daily work.

Managers’ responsibility is to own and manage the context, the limitations and triggers of improvement as well as developing an understanding of the relationships between the parts of the system. This originates from the findings that altering on the job behavior goes through focusing on task alignment as one of the core ideas of PCTA.

**Principle 2: Minimize risk - locally adapt improvement ideas by tests on a limited scale**

This principle is to acknowledge the high complexity and low linearity of care thus minimizing effects from failures by testing how ideas can be adapted to a local context on a limited scale (paper III and IV). From a learning perspective the idea of small scale testing to adapt ideas to local context provides an opportunity to discover the response to a specific change within the system, which is consistent with the need for local adoption concluded in paper IV.

This strategy also provides managers with the opportunity to loosen management control – since the stakes are low. It is thus a way to empower staff and nurture them into the understanding and joy of improving their own situation.

**Principle 3: Create feedback systems based on performance measurement**

As shown in paper II, only small amounts of data were readily available for use in those units where the ICS strategy was employed in improvement work. Thus it is reasonable to presume that there were also few formulated goals with respect to using data. However, it is at the core of the PCTA strategy to use data and set goals as well as to focus on managerial support systems (paper III). Such systems should pay attention to the needs of data that enable practitioners and managers to conduct changes in on the job behavior from a strategic patient perspective i.e. to support the PCTA strategy.

**Principle 4: Create arenas for organization wide strategic agreements**

Closely related to goal settings and interpretation of data is the profound idea of social constructs supported in paper III. The PCTA strategy puts effort into developing arenas for discussions that can create shared social constructs about the aims and performance of the organization. I argue that the establishment of these arenas and the choice of topics for discussion are one of the most important managerial tasks. Here it is possible to link healthcare practice to management intentions, based on data.

Together, the above principles may serve as the starting point for managerial teams to decide how to set up their own arenas to continuously develop and refine the application of teleology as described through the PCTA strategy. As stated in the introduction, the
possibilities of attaining goals are as likely as finding the pot of gold at the end of the rainbow. Since goals are social constructs, there is a continuous need to move on and use arenas to discuss and decide upon alternatives to current patient centered task aligned actions.

**Principle 5: Create an environment for policymaker’s continuous learning**

In the paragraph *Managers do their best* it is argued that there is a general lack of knowledge about how to manage a PCTA strategy. An environment for managers’ and policymakers’ continuous learning is therefore important.

I argue that top managers need arenas for reflections upon and learning from change initiatives in order to understand more about effects from changes in their systems in relation to their own management actions. Also, policymakers need to create the foundation for the PCTA strategy together. This implies sharing ideas about the meaning of the five principles within their own context as well as improving their own abilities to manage the improvement process.

The famous psychologist Kurt Lewin once proposed that: “if you think you understand the system, try changing it…” The point made is that we can not really understand our systems by just observing the status quo. The causality between management’s decisions and activities in practice are not so clear and straight forward. The strategic learning process in which top managers participate can be one arena in which they learn about and reflect upon relations between their endeavors and the events occurring in change processes. Some of the requirement for this process to take place is that sufficient time is created for attention towards improvement and change.
Adaptation of principles

In order to get the PCTA strategy to work in practice it is probably good to start developing the new strategy based on some easy successes that are aligned with the company wide strategy. Generating some fast positive feedback gives both practitioners as well as managers courage to carry on. However, it is important not to only pick the easy improvements since such strategy might lead to an extensive activity that does not necessarily lead to improved systems outcomes. From this standpoint there is a trade off between results from a short and long perspective.

The idea of PCTA is neither a quick fix nor a magic bullet – it is hard practical work. Management needs to develop a common set of core ideas to support their long term strategy. They also need to create a climate which stands for constancy of purpose for the new strategy. When the work gets tough it is important not to fall back to old behaviors, especially in the beginning of the adoption of the new strategy since forces from the old system will rise.

Finally, the transferability of these principles into action is contextually dependent. This puts emphasis on the need for local adaptation. The process of local adoption relies both in the capability of building knowledge of the local care system as well as a profound knowledge of how to go about change and improvement. The acquisition of these required knowledge bases are both a question of theories as well as a question of action. Schön (Schön 1983) describes the properties of the intellectual process embedded in the needed knowledge inquiry as “When someone reflects-in-action… He does not keep means and ends separate, but defines them interactially as he frames a problematic situation… Because his experimenting is a kind of action, implementation is built into his inquiry”.

By these words I want to encourage managers interested in pursuing teleology based on the PCTA strategy to start experimenting in order to learn what their organizations best ways are to apply these empirically based principles ∞Δ
References


Original papers I – IV