

“ In God we trust.  
All others, bring data.”  
W. E. Deming (Out of Crisis)

**Maj juli 2010**  
**Kjeld Møller Pedersen &**  
**Jacob Anhøj**

## **Evaluering af ’Patientsikkert Sygehus’**

### **Indholdsfortegnelse**

Terminologi.....	2
Evaluering af komplekse interventioner i QI-sammenhæng.....	3
Og hvad skal man så konkludere?.....	6
Baggrund 1 for Patientsikkert sygehuse: Formål m.m.....	7
Baggrund 2: Ønsker til evaluering .....	7
Baggrund 3: Målinger og målestrategi.....	8
Overordnede målsætninger .....	8
Målestrategi og -metoder .....	8
Identifikation og beskrivelse af forbedringer.....	10
Baggrund 4: Refleksioner over målstrategi, jfr. ’Baggrund 3’ .....	10
Problemstillinger ved denne evalueringsstrategi: .....	11
To mulige evaluerings-strategier .....	12
Bruge afdelinger/afsnit på pilothospitalerne som sammenligningsgrundlag.....	12
Evaluering med sammenligningshospitaler/afsnit, hvor pakkerne ikke er indført .....	13
Økonomisk evaluering (omkostnings-effekt) .....	13
Litteratur.....	15

Dette notat drejer sig om mulighederne for at evaluere/effektmåle initiativet 'Patientsikkert Sygehus'. Indledningsvis skal bemærkes, at det altid er svært at udforme en strategi for effektmåling, efter at et projekt er sat i gang. Overvejelser om dette bør være en integreret del af forberedelsen til et projekt. Derfor vil det følgende i betydelig grad være præget af, hvad der forekommer realistisk i den foreliggende situation og ikke, hvad der mere ideelt kunne være gjort, hvis det havde været indtænkt som en integreret del af projektet.

For at forstå begrundelsen for den foreslåede evaluerings-strategi er det nødvendigt med en del baggrund, fordi Patientsikkert Sygehus er baseret på en projektfilosofi og evalueringstænkning, som der langt fra er enighed om i litteraturen og blandt praktikere – og den foreslåede evaluerings-strategi følger kun i beskedent omfang denne tænkning. – Den utålmodige læser kan springe dette over og gå direkte til det konkrete forslag, p. 8/9, med fare for at komme til at mangle nogle mellemregninger.

## ***Terminologi***

Indledningsvis lidt begrebsafklaring, bl.a. for at danne baggrund for den evalueringsdiskussion, der pågår omkring projekter med kvalitetsudvikling som hovedformål. Begrebet **evaluering** dækker kort sagt over 'bedømmelse', det vil sige, at man fastsætter/vurderer værdien af en given aktivitet. I den forstand er evaluering dermed en aktivitet, som foregår hele tiden som en mere eller almindelig hverdagsaktivitet. Videnskabeligt dækker 'evaluering' over et sæt af metoder/værktøjer, som kan bruges til mere eller mindre systematisk at evaluere en aktivitet. Der kan anlægges (mindst) tre forskellige videnskabsparadigmer: En positivistisk, en hermeneutisk og en konstruktivistisk. I nærværende sammenhæng er det primært den første tilgang, der udgør forståelsesrammen.

Der er to hovedformer for evaluering: '**Resultat-evaluering**' – også her kaldet **effekt-måling** - og '**proces-evaluering**'. De besvarer to forskellige, men ofte komplementære, spørgsmål. I effektmålingen er hovedopgaven at besvare følgende spørgsmål: Førte aktiviteten til det ønskede resultat (mål) – og kan man være sikker på, at en given grad af målopfyldelse skyldes aktiviteten (kausalitet), medens man i proces-evalueringen besvarer spørgsmålet om, 'hvad skete det', bl.a. hvordan blev aktiviteten gennemført, kontekstuelle omstændigheder m.m.. (alternative begreber er hhv. 'summativ evaluering' og 'formativ evaluering')

Ved **design** forstås de undersøgelsesmæssige rammer for evalueringen. Et design kan være være eksperimentelt (randomiseret) eller quasi-eksperimentelt (3;4), som omfatter alle andre design-former end det randomiserede, både med og uden kontrolgruppe, incl. case-control. I epidemiologien arbejdes endvidere med observationelle studier, bl.a. kohorter. Designs vurderes især ud fra deres interne validitet, dvs. fravær af bias og dermed evnen til (nogenlunde) éntydigt at kunne drage kausale konklusioner af typen: "en observeret ændring 'skyldes' interventionen..

Design og type af **intervention** hænger sammen. Der kan groft taget skelnes mellem **'simple'** og **'komplekse' interventioner** – som samtidig fører over i en evaluerings-uenighed i forbindelse med 'quality improvement', **QI**, og Patientsikkert Sygehus kan netop klassificeres som **QI**.

QI omfatter bl.a. 'forbedrings-og forandringsinitiativer', se tekstramme. Nogle typer af designs er mere vel-egende til simple interventioner, typisk randomiserede designs, end til komplekse interventioner, hvor der ofte er tale om quasi-eksperimentelle designs – eller, som det fremgår senere, at man ikke mener, at evaluering indenfor de etablerede forståelsesrammer er mulig, fordi der er tale om praksisnære forandringsprojekter. I det følgende gennemgås og diskuteres denne uenighed.

The group **defined QI** as systematic, data-guided activities designed to bring about immediate improvements in health care delivery in particular settings. Quality improvement uses an array of methods and can look like practical problem solving, an evidence-based management style, or an application of a theory-driven science of system change. Quality improvement methods often encourage health care workers to use their experience, along with insights from others, to identify promising improvements, implement changes on a small scale, monitor and interpret effects, and decide about additional changes and wider implementation. ...  
At its heart, QI is a form of *experiential learning* that regards improvement to be part of the work process and always involves deliberate actions expected to improve care, guided by *data reflecting the effects*. (1) – KMPs fremhævninger

### ***Evaluering af komplekse interventioner i QI-sammenhæng***

Det følgende er vigtigt, fordi det er baggrunden for og en illustration af den opfattelse af evaluering og QI, som IHI står for – IHI, som har leveret konceptet for Patientsikkert Sygehus.

Det klassiske eksempel på en '*simple intervention*' er et lægemiddelforsøg med to piller/præparater, jfr. det af Bradford Hill ledede klassiske randomiserede forsøg: "Medical Research Council Streptomycin in Tuberculosis Trials Committee: Streptomycin treatment of pulmonary Tuberculosis"., BMJ 1948; ii: 769-782. Dette forsøg betragtes som det første egentlige randomiserede forsøg i lægevidenskaben og er forbilledet for designs for simple interventioner. Interventionen er simpel (medicinen), og man har fjernet enhver organisatorisk kontekst og mekanisme ved hjælp af randomiseringen, dvs. designet har høj intern validitet. Man har isoleret effekten af pillen og kan med rimelig sikkerhed tale om en kausal sammenhæng. Den komplekse intervention på den anden side medtænker den organisatoriske dimension – og organisatoriske ændringer er en del af interventionen.

Campbell et al (5;6) beskriver problemstillingerne omkring *komplekse interventioner* på følgende måde – svarende ganske godt til pakkerne og deres organisatoriske kontekst i Patientsikkert Sygehus (med mine fremhævninger):

"Complex interventions are "built up from a number of components, which may act both independently and interdependently." *Many health service activities should be considered as complex. Evaluating complex interventions can pose a considerable challenge* and requires a substantial investment of time. *Unless the trials illuminate processes and mechanisms they*

*often fail to provide useful information. If the result is negative, we are left wondering whether the intervention is inherently ineffective (either because the intervention was inadequately developed or because all similar interventions are ineffective), whether it was inadequately applied or applied in an inappropriate context, or whether the trial used an inappropriate design, comparison groups or outcomes. If there is a positive effect, it can be hard to judge how the results of the trial might be applied to a different context”*

Der er en livlig diskussion om, hvad der er den ‘rigtige’ måde at evaluere komplekse interventioner på, bl.a. fordi sådanne interventioner ofte *ikke* har en stringent protokol med en éntydig definition af interventionen og foregår i et praksis-og forbedringsorienteret miljø, hvor interventionen løbende ændrer karakter i modsætning til den simple intervention.

Af citatet kan man som minimum se, at man ved komplekse interventioner skal have en proces-evaluering for at dokumentere, hvad der er foregået (nok som supplement til en resultat-evaluering). Inden for randomiserede forsøg har praksisfokuseringen ført til den design-kategori, der betegnes som ‘pragmatic’ trials (7;8), bl.a. med klyngerandomisering og randomisering af hele organisationer i modsætning til randomisering af individer.

En lidt *anden dimension af komplekse interventioner* finder man i **(små/mindre) projekter**, der sigter mod at forbedre/udvikle kvalitet eller patientsikkerhed, ofte på relativ kort tid. I en vis forstand kan man opfatte hver af pakkerne i Patientsikkert Sygehus som sådan et mindre projekt. Det er her, hvor metoder som Gennembrudsmetoden, PDSA-cirklen/Model of Improvement, statistisk proceskontrol m.m. ofte bringes i spil. Davidoff m. fl. (9;10) har i en række artikler argumenteret for, at der bør opbygges retningslinjer for rapportering fra kvalitetsudviklings-projekter, (quality improvement, QI). Det skyldes, at der publiceres alt for lidt vedrørende resultaterne af lokale QI-projekter, bl.a. fordi de ud fra standard-kriterier kun i mindre grad lever op til almindelige kriterier for videnskabelig afrapportering.

Davidoff m. fl. noter bl.a. (med mine fremhævninger; Davidoff er executive editor på IHI)

“Unlike conceptually neat and procedurally unambiguous interventions such as drugs, tests, and procedures that directly affect the biology of disease, and are the objects of study in most clinical research, *improvement* is essentially a social process. *Improvement is an applied science rather than an academic discipline; its immediate purpose is to change human performance, rather than generate new, generalisable knowledge, and it is driven primarily by experiential learning.* Like other social processes, improvement is inherently context dependent; it is reflexive, meaning that *improvement interventions are repeatedly modified in response to outcome feedback, with the result that both its interventions and outcomes are relatively unstable; and it generally involves complex, multicomponent interventions.* Although traditional experimental and quasi-experimental methods are important for learning whether improvement interventions change behaviour, they do not provide appropriate and effective methods for addressing the crucial pragmatic (or “realist”) questions about improvement that are derived from its complex social nature: what is it about the

mechanism of a particular intervention that works, for whom, and under what circumstances?”, (10)

Davidoff og medforfatterne drager således den slutning, at ikke blot kan man ikke bruge traditionelle designs (både randomiserede og quasi-eksperimentelle) men siger i realiteten, at evalueringen primært skal være en proces-evaluering, jfr. de tre spørgsmål sidst i citatet ('what is it', 'for whom', 'under what circumstances') og ser således tilsyneladende væk fra hovedeffekten af interventionen. Samtidig indføres en skelnen mellem 'applied science', dvs. det praksisnære, som Berwick og andre også betegner som 'pragmatic science', (11), og 'academic discipline', dvs. det traditionelle videnskabelige (evaluering)paradigme. Denne skelnen benyttes således som argumentation for to forskellige evalueringsoptagelser, omend de konkrete retningslinjer der foreslås i SQUIRE-

Berwicks illustration og argumentation vedr. pragmatic science (en ny epistemologi).

The argument for that epistemology is not a simple one, but its intuitive force is somewhat easier to uncover with a simple question: "How much of the knowledge that you use in your successful negotiation of daily life did you acquire from formal scientific investigation—yours or someone else's?" ...

Broadly framed, much of human learning relies wisely on effective approaches to problem solving, learning, growth, and development that are different from the types of formal science so well explicated and defended by the scions of evidence-based medicine. Although they are far from RCTs in design, some of those approaches offer good defences against misinterpretation, bias, and confounding. In the world of clinical care, especially in the quest for improvement of clinical processes, is it plausible that those approaches—the ones we use in everyday life—might have value too, used well and consciously, to help us learn?

rappporteringsformatet, ikke er helt så unuancerede, (10), som debatten antyder, utvivlsomt fordi der har været tale om en konsensus-proces, hvor andre end IHI-relaterede personer har medvirket.

I forlængelse af tænkningen bag 'pragmatic science' ser man også en opfattelse af, at der ikke behøves de samme data (eller samme datakvalitet?), som ved mere traditionelle akademiske evalueringer.. Data afhænger af, lyder argumentatione, om der måles af hensyn til forskning, forbedring eller bogholderi/sammenligning?, (12). Modsynspunktet går på, at der ikke som sådan er forskel, og at man skal følge de almindelige krav til datakvalitet, (13). I sidstnævnte artikel opstilles følgende tabel som en slags checkliste for krav til datakvalitet i QI-projekter.

Table 1 Data quality control methods for QI projects

Project phase	Challenge question
Project design	Are the aims of the project clearly stated? Is a valid definition and measurement system available for the required data? Is there a clear focus on quality, rather than quantity, of data?
Data collection	Is a standardized data collection form created? Are data items clearly defined and written instructions provided for collecting each data item? Are staff adequately trained to collect data? Are QA reviews completed? Is an electronic database used for data management? Are sufficient database controls in place to identify errors? Is there a back-up routine for the electronic database?
Data management	Have data been evaluated using basic statistics? Has there been a comprehensive review for missing data and methods to minimize missing data?
Data analysis	Are missing data reported and appropriate methods used to account for it? Have potential outliers been identified and evaluated? Have appropriate methods been used to provide summary measures of the project results? Have measures of precision been presented with the study results? Have appropriate methods been used to evaluate the impact of factors that may confound the results?

Dette ser man tydeligt i den debat, der udspandt sig omkring Davidoffs første artikel. Temaet var, om praxisfor-ankrede kvalitetsforbedringsprojekter kunne/ikke kunne evalueres på mere traditionel vis. Peter Provonost betegnede det som et skridt frem og to tilbage, (13). Grundsynspunktet var, at

“**T1 translational research** as “the transfer of new understandings of disease mechanisms gained in the laboratory into the development of new methods for diagnosis, therapy, and prevention and their first testing in humans” ... and **T2 translational research** as “the translation of results from clinical studies into everyday clinical practice and health decision making” Westfall and colleagues<sup>11</sup> have proposed an additional step, **T3 translational research**, as “practice-based research to translate distilled knowledge from guidelines and systematic reviews to day-to-day clinical care” ... The journey from science to action (i.e. T2 and T3) can take decades. *Accelerating* these translational research steps requires interaction between and collaboration among different skills and disciplines .... Furthermore, funding for T2 and T3 translational research by various agencies needs to be clarified and prioritized” (KMPs fremhævninger), (2)

‘many QI studies have significant bias and have a high risk of misinforming readers’ (13).

En lidt anden italesættelse finder man hos Shojania, Grimshaw m.fl., (2;14-18). Her er nøgleordene *translational research og implementering*.

I sin grundform drejer translational research sig om at om-

sætte forskningsresultater (mere grundlæggende, eller fx. kliniske retningslinjer) til praksis. Dette indebærer ofte ændrede rutiner, ændret kultur, nye læringsmodeller osv..

Grundlæggende kan *Patientsikkeret Sygehus opfattes som translational eller implementeringsforskning*. Afsættet er evidensbårne rutiner, som skal spredes. Modellen bag Patientsikkeret sygehus er således, at man bl.a. giver rådgivningsbistand til de involverede afdelinger, men overlader også meget til lokale initiativer/udformning af ændringsprocessen. Om pakkerne hedder det bl.a. ”Filosofien bag de kliniske pakker er at samle en række interventioner, der hver især er udtryk for *best practice* og formodes samlet set at give et bedre klinisk resultat”

## Og hvad skal man så konkludere?

- der er naturligvis forskel på simple og komplekse interventioner – og evalueringsdesignet skal afspejle dette. Pakkerne i Patientsikkert Sygehus er hver for sig en kompleks intervention og tilsammen måske ’superkomplekst’.
- der er naturligvis forskel på at afprøve under drivhus-forhold, svarende til hvad der ofte er tale om med standardiserede protokoller ved simple interventioner, og afprøvning under de forhold, som en almindelig køkkenhave giver. Grundiddeen med Patientsikkert Sygehus er implementering i en dagligdagssituation, men hvor man ikke kan fortsætte med ’business as usual’. Der kræves ændrede rutiner, ændret kultur m.m. – som forudsætning for at kunne forvente en bedre klinisk effekt end under de ændrede rutiner.
  - Dette skal dokumenteres. Der kræves med andre ord en dokumentation for, hvordan/på hvilken måde/under hvilke forudsætningerne blev implementeret
    - er dette ikke tilfældet kan man jo ikke på dokumentet vis sprede evt. positive erfaringer.
    - dette betyder, at der skal indgå en proces-evaluering (i tillæg til en resultat-evaluering.)
- implementering af behandlinger/pakker, som har en kliniske effekt, som er demonstreret i gode kliniske forsøg, kan implementeres på en sådan måde, at den ønskede effekt måske

udebliver, fordi implementering i en anden organisatorisk kontekst, en anden (sikkerhedskultur) osv. kan påvirke effekten

- et argument om, at man implementerer for at lære ændrer ikke ved, at man fortsat skal kunne dokumentere effekt
- derfor bør man også ved (mindre) lokale forbedringsprojekter sikre at gængse regler fra den videnskabelige evalueringslitteratur følges.
- især, hvis man mener, at erfaringer fra vellykkede lokale projekter skal spredes til andre afdelinger/institutioner. Dette kræver erfaringsopsamling og erfaring fra flere mere eller mindre sammenlignelige projekter.
  - hvilket bl.a. stiller krav til datakvalitet/validitet

### ***Baggrund 1 for Patientsikkert sygehus: Formål m.m.***

Det følgende er i al væsentlighed baseret på oplysninger på hjemmesiden, [www.sikkerpatient.dk/patientsikkertsygehus.aspx](http://www.sikkerpatient.dk/patientsikkertsygehus.aspx)

Patientsikkert Sygehus har til formål at forebygge utilsigtede fejl, skader og dødsfald ved at man på fem pilotsygehuse implementerer 12 pakker. Implementeringen understøttes på forskellig vis. Projektperiode er fra 2010 til 2013

Ambitionen er at opnå 15 % reduktion i dødeligheden og 30 % reduktion i skader ved fx at reducere antallet af hjertestop, eliminere en række hospitalsinfektioner, reducere forekomsten af tryksår og forebygge medicineringsfejl. De gode resultater skal spredes til og være inspiration for landets øvrige sygehuse

En væsentlig drivkraft bliver, at både sygehuse og de enkelte afdelinger vil sætte konkrete og ambitiøse mål, som man vil stræbe efter at nå i løbet af projektet. Fx konkrete mål for reduktion af hjertestop, sygehusinfektioner og medicineringsfejl.

Patientsikkert Sygehus arbejder med forbedringsmodellen, Model for Improvement, der er udviklet af Institute for Healthcare Improvement, IHI.

Forbedringsmodellen er et enkelt og meget anvendeligt værktøj til at accelerere forandrings- og forbedringsprocesser..

### ***Baggrund 2: Ønsker til evaluering***

Nærværende notat er udarbejdet i forbindelse med arbejdet i referencegruppen vedr. evaluering. Opdraget til referencegruppen lyder (med mine fremhævninger):

”Arbejdsgruppen bidrager med forslag til, hvordan projektet kan evalueres. Der lægges vægt på målbare indikatorer, som repræsenterer forskellige sider af projektet, fx økonomi (hvad koster indsatsen?, hvad er udbyttet?), kultur (hvad betyder deltagelse for patientsikkerhedskulturen på de deltagende sygehuse?) og klinisk effekt (er dødelighed og skadesforekomst reduceret på de deltagende sygehuse?). Barrierer og muligheder for implementering (hvad skal der til for at gennemføre vedvarende forbedringer i klinikken”

### **Baggrund 3: Målinger og målestrategi**

Nedenstående er uddrag fra et selvstændigt dokument om målinger og målestrategi. I dokumentet findes desuden en samlet oversigt over alle indikatorer med tilhørende operationelle definitioner.

Den overordnede målsætning er at reducere dødsfald målt med hospitalsstandardiseret mortalitetsrate (HSMR) med mindst 15% og forekomsten af skader målt med Global Trigger Tool med mindst 30% på de fem projektsygehuse.

For at opfylde disse målsætninger skal projektsygehuse implementere 12 kliniske ”pakker”, som hver for sig bidrager til reduktion af død og skader.

For hver pakke er der udarbejdet en eller flere procesindikatorer til løbende at måle pakkens implementeringsgrad. Til de pakker, hvor det har været muligt, findes desuden en resultatindikator til at dokumentere, at implementeringen har den ønskede virkning for patienterne, fx færre hjertestop.

Med baggrund i danske og internationale erfaringer fra sygehuse, som har opnået gode resultater opstilles der så vidt muligt målsætninger for hver resultatindikator. Fx ved vi i dag, at det er muligt stort set at eliminere visse hospitalserhvervede infektioner, som tidligere blev betragtet som uundgåelige komplikationer. Og fra sygehuse, som har opnået disse og andre forbedringer, ved vi, at det er muligt inden for en overskuelig periode at reducere dødsfald og skader.

#### **Overordnede målsætninger**

- Den hospitalsstandardiserede mortalitetsrate (HSMR) reduceres med mindst 15%
- Antal patientskader per 1000 sengedage målt med Global Trigger Tool (GTT) reduceres med mindst 30%
- Antal hjertestopkald reduceres med mindst 30%
- CVK-relaterede infektioner elimineres
- Respiratorrelaterede lungeinfektioner elimineres
- Andelen af patienter, som udvikler tryksår under indlæggelse, reduceres med mindst 50%
- Andelen af patienter med nydiagnosticeret hjertesvigt, som genindlægges inden for 28 dage, er højst 10%
- Andelen af opererede patienter, som dør under indlæggelse, reduceres med mindst 20%
- Andelen af opererede patienter, som genindlægges inden for 30 dage, reduceres med mindst 20%

#### **Målestrategi og -metoder**

Målestrategien i dette projekt består af følgende skridt:

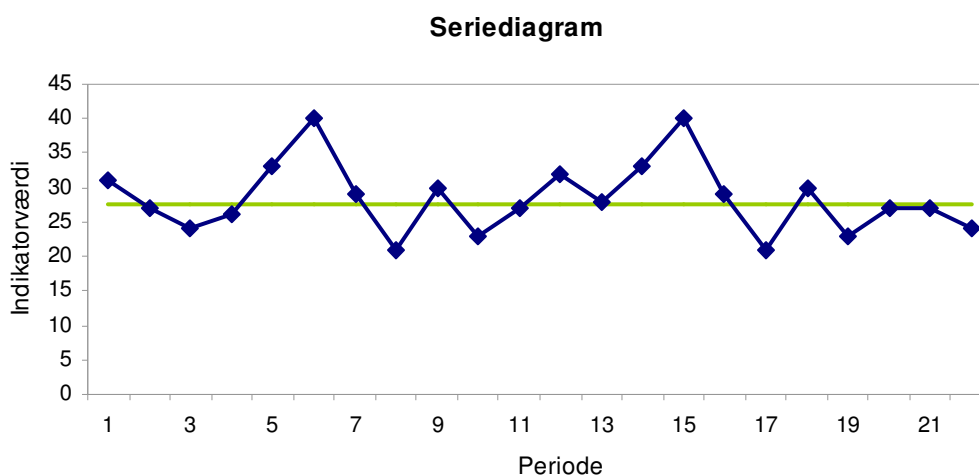
1. Udvælge proces- og resultatindikatorer
2. Udarbejde præcis operationel definition for hver indikator
3. Beskrive plan for dataindsamling
4. Foretage en statistisk analyse af hver indikator for at afgøre, om der sker ændringer

## 5. Koble analyseresultater til forbedringsstrategier

Forbedring i en proces kan udtrykkes enten som et absolut mål (fx nul respiratorrelaterede lungeinfektioner eller 100% overholdelse af håndhygiejneprocedurer) eller som en relativ ændring (fx mindst 30% reduktion af hjertestopkald).

Om der opnås forbedringer, afgøres i forhold til baseline, som er seneste stabile periode før eller umiddelbart efter projektstart. Stabilitet bedømmes så vidt muligt på baggrund af mindst 12 (helst 15 til 20) på hinanden følgende målepunkter, som kun udviser tilfældig variation.

Ikke tilfældig variation identificeres ved hjælp af et seriediagram, som viser indikatoren på en kurve med tiden på x-aksen. Midt i diagrammet indtegnes medianen som en vandret linje.



Før seriediagrammet fortolkes, opgøres antallet af serier. En serie er et eller flere på hinanden følgende målepunkter på samme side af midtlinjen.

Herefter benyttes fire regler til identifikation af mønstre i data, som tyder på ikke tilfældig variation:

1. **Skift:** Seks eller flere på hinanden følgende målepunkter på samme side af medianen (punkter på medianen ignoreres)
2. **Trend:** Fem eller flere på hinanden følgende målepunkter som stiger eller falder (punkter med samme værdi som det foregående punkt ignoreres)
3. **For få eller for mange serier** i forhold til antallet af målinger (findes ved opslag i standardtabeller)
4. Tydeligt **afvigende målinger** (dette er subjektiv afgørelse)

Ikke tilfældig variation kan opstå af to primære årsager:

1. En udefrakommende eller uventet faktor har påvirket processen.
2. Der er (med vilje) introduceret forandringer med det formål at påvirke processen i den ønskede retning.

I begge tilfælde tyder ikke tilfældig variation på, at processen, som indikatoren måler, er under forandring. Variation i den ønskede retning, anses som forbedring. Ikke tilfældige variationer, som ikke kan tilskrives forbedringsstrategier, bør på den anden side altid undersøges nærmere for at fastslå årsagen til forandringen.

Ideelt set bør der foreligge data fra et år forud for påbegyndelse af forbedringsarbejde for hver enkelt indikator. På denne måde har man et tydeligt billede af, hvordan indikatormålingerne varierer over tid, og af hvilken type af variation, som de forskellige processer indeholder. I mange tilfælde er det imidlertid ikke muligt at skaffe historiske data for indikatorerne, fordi disse ikke tidligere er blevet målt. I disse tilfælde, tages alternative metoder i brug til at fastlægge baseline.

For det første vil det i mange tilfælde, afhængig af mængden og tilgængeligheden af data, være muligt at foretage indikatormålinger hyppigere end månedligt, fx ugentligt eller ligefrem dagligt. Dette gør det muligt at indsamle baselinemålinger hurtigere samtidig med, at forbedringsarbejdet foregår.

For det andet kan man for procesindikatorerne forvente at se hurtige forandringer (= ikke tilfældig variation i ønsket retning), hvis forbedringsarbejdet har de ønskede effekter. Resultatindikatorerne derimod vil forventeligt først udvise forandringer senere i projektet, når de underliggende processer er under kontrol. Dermed er det muligt at foretage en pålidelig beskrivelse af baseline på resultatindikatorerne, også selv om, der ikke foreligger historiske indikatormålinger.

Som udgangspunkt opgøres alle indikatorer på det niveau, målingen sker på. Mortalitet, hjertestopkald og patientskader opgøres således på sygehusniveau, mens fx tryksår opgøres på afdelingsniveau.

## **Identifikation og beskrivelse af forbedringer**

Forbedring kan være forbigående eller permanent. For at en opnået forbedring kan anses for permanent, skal processen indstille sig på et nyt stabilt og bedre niveau end baseline. I projektperioden foretages løbende rapportering af indikatormålinger fra sygehusene. Målingerne vurderes og kommenteres løbende af IHI's og DSFP's måleeksperter.

Slutniveauet bestemmes som den seneste stabile periode før eller umiddelbart efter projektets afslutning ultimo 2012.

Den opnåede forbedring er således forholdet eller forskellen mellem medianen ved den seneste stabile periode og ved baseline. Yderligere benyttes de fire regler for identifikation af ikke tilfældig variation i seriediagrammer for at sikre, at indikatormålingerne ved projektets afslutning kun udviser tilfældig og forudsigelig variation.

For indikatorer, hvor målet er nul hændelser (fx CVK-relaterede infektioner), anses målsætningen for opnået, hvis der ingen hændelser har været i mindst 300 dage.

## ***Baggrund 4: Refleksioner over målstrategi, jfr. 'Baggrund 3'***

**Designmæssigt** er målestrategien udtryk for

- før – efter-måling (uden kontrol/sammenligningsgruppe)
  - ved 'før' forstås den baseline-måling der omtales i (19)
  - hver pakke på hver sygehus tænkes evalueret på denne måde.
    - der er således tale om en evaluerings-strategi hvor hver enkel intervention undersøges for sig – og ingen tværgående sammenligninger. I sagens natur giver det betydelige problemer med den interne validitet.

- Der er tale om både proces- og resultat-evaluering/effektmåling

**Effekt mål:** Fremgår af nedenstående tabel for de enkelte pakker, incl. registreringsenhed Ved en række af pakker er der indbygget en slags 'kausal-kæde'. AMI-pakken kan betragtes som eksempel:

- Procent AMI-patienter, som modtager alle elementer i AMI-pakken (proces-indikator)
  - > % AMI-pt, som dør under indlæggelse
    - det er åbenbart at en nødvendig forudsætning for at en evt. observeret ændring i dødeligheden kan tilskrives pakken er, at (noget eller hele) pakken har været en del af behandlingen

På sygehusniveau tænkes anvendt hospitalsstandardiseret mortalitets rate, HSMR. Det baseres på centralt foretagne beregninger.

**Datamæssigt** er der tale om

- lokalt indsamlede og lokalt lagrede data, som tænkes bearbejdet i serie-diagrammer, jfr. tabellen på næste side m. resultat/proces-indikatorer.
- data tænkes *ikke* 'standardiseret', dvs. baseret på fælles og klare definitioner som bruges af de fem pilohospitaler således at der kan sammenlignes på tværs af afdelinger/hospitaler data tænkes *tilsyneladende ikke* samlet nationalt med henblik på sammenligning på tværs af afdelinger/sygehuse
- på nuværende tidspunkt kan det ikke vurderes om dataindsamlingen kvalitetssikres, fx. for *fuldstændighed* (er alle patienter registeret)
- Det kan ikke afgøres om der er foretaget styrkeberegninger. Er der en berettiget forventning om, at der indgår patienter nok til statistisk (nogenlunde) sikkert at kunne dokumentere en ændring (ud over det visuelle i serie-diagrammet)

### **Problemstillinger ved denne evalueringsstrategi:**

- den interne validitet af en før-efter-sammenligning er ikke overbevisende
- i den foreliggende form har erfaringer stort set kun værdi lokalt (bl.a. på grund af manglende sammenligning med andre enheder)
  - det bliver svært på dokumentret vis at sprede gode erfaringer, fordi der (tilsyneladende) ikke indgår en proces-evaluering, hvor man dokumenterer den præcise indsats)
  - der er i realiteten ikke stor ekstern validitet
- Brug af HSMR på hospitalsniveau kan diskuteres, både ud fra HSMR-tænkningen og især ud fra spørgsmålet om internvaliditet, (20;21)

## ***To mulige evaluerings-strategier***

Det følgende skal opfattes som mulige strategier, hvor en del af den skitserede målestrategi kan genbruges under visse forudsætninger, bl.a. Global Trigger Tool. Det centrale er i begge de skitserede strategier forsøger der indarbejdet sammenligningsgrupper. Grundproblemet med såvel den lokale 'før-efter' strategi er, at man aldrig kan vide, om en (stabil) ændring med rimelighed kan henføres til interventionen eller er udtryk for en generel tendens som i (varierende grad) også kan observeres på sygehuse, som ikke deltager i projektet.

### **1. bruge afdelinger/afsnit på pilothospitalerne som sammenligningsgrundlag for hinanden**

- baseret på den antagelse, at pakkerne er implementeret forskelligt
- forudsætter såvel standardiserede data, jfr. den indsatte tabel foran, som at data gøres tilgængelige centralt for fælles analyse
  - i. bruger de i nedenstående tabel anførte effektmål, jfr. også målstrategien
  - ii. overveje ikke at bruge HSMR
  - iii. bruge 'global trigger tool' til at estimere den relative ændring i forekomsten af patientskader på sygehusniveau.
- at der gennemføres *en proces-evaluering*, fx. på tre tidspunkter, hvor karakteren, omfang og kontekst af implementering af pakkerne dokumenteres
  - i. den præcise karakter af process-evalueringen skal uddybes. Der findes ikke som sådan en fast etableret metodik. En del af dokumentationen vil evt. kunne indsamles via selv-registreringer, medens andre forhold ('kontekst') skal indsamles mere systematisk ('særlige lokale forhold, fx. mange 'ildsjæle'; allerede etablerede rutiner, organisatoriske forhold osv..).
- survey blandt relevante personalegrupper (de berørte sengeafd./intensiv osv.) omkring sikkerhedskultur. Surveyen skal gentages mindst to gange, og gerne tre (start, midtvejs og ved slutningen).

Global-trigger-tool, GTT, (19;22-26), kan bruges til at estimere antal skader (adverse events) på sygehusniveau. Hvis alle 12 pakker indføres systematisk på et sygehus, er det rimeligt at antage, at det vil påvirke antal skader. Grundideen i IHIs GTT er, at man hver 14. dage ved en tilfældig stikprøve udtager 10 journaler, dvs. omkring 250 journaler/år, som gennemgås for at identificere 'triggers', dvs. markører for mulig skade. Tilstedeværelsen af en trigger udløser en nærmere granskning af journalen mhp. at identificere eventuelle patientskader. De identificerede skader beskrives og klassificeres efter skadens alvor. Hver journalgennemgang tager højst 20 minutter, i gennemsnit ca. 10. Gennemgangen foretages uafhængigt af to erfarne sygeplejersker, efterfulgt af en afstemning med en supervisor, som altid er læge. Metodens sensitivitet angives til ca. 80% sammenlignet med en "guld-standard-journalaudit". Specificiteten er meget høj (~100%). Interobservatøroverensstemmelsen er høj ( $\kappa > 0,8$ ) for trænedede reviewere (Classen. et al. 2008)

Global Trigger Tool har bl.a. været afprøvet i Region Syddanmark, <http://www.centerforkvalitet.dk/wm293462>, og i Region Hovedstaden.

## **Evaluering med sammenligningshospitaler/afsnit, hvor pakkerne ikke er indført**

- giver større intern (og ekstern) validitet
- et matchende sygehus per pilothospital/afdeling
  - i. skulle ikke være en umulig opgave
- samme effekt-mål som i nedenstående tabel
  - i. usikkerheden er, om der (nogle) af disse forhold allerede registreres på andre sygehuse – og hvis ikke, hvad er krævet af incitamenter for at det kan gøres
- Bruge Global Trigger Tool (der også vil være det absolut minimale, der kan indsamles fra sammenligningssygehuse).
- proces-evaluering, som beskrevet ved model 1
- survey om sikkerhedskultur, jfr. model 1.
- Usikkerheden er, om det er rimeligt at antage, at der findes sygehuse i Danmark, som ikke på den ene eller anden måde arbejder med indholdet i pakkerne – altså, er det muligt at isolere effekten til de fem deltager-sygehuse?

## **Økonomisk evaluering (omkostnings-effekt)**

- Ved økonomisk evaluering forstås her omkostnings-effekt-analyse. Som navnet antyder forudsætter det både oplysninger om effekten af en intervention og (netto-omkostningerne) ved indsatsen.
- for at give mening forudsætter det enten model 1 eller 2 (ideelt set model 2)
- der skal etableres dataindsamling vedr. ressourceforbrug (både midertidigt i forbindelse med opstart og mere permanent)
- Muligvis vil der blive tale om en 'omkostnings-minimeringsanalyse, hvis der ikke kan findes et fælles effektmål for pakkerne. Et muligt fælles-effekt-udtryk kunne dog evt. dannes ud fra opgørelser af Global Trigger Tool og så beregne netto-omkostninger per pct. ændringer i adverse events. Ved netto-omkostninger forstås her driftsomkostninger minus evt. besparelser ved et fald i antal adverse events. Er dette ikke muligt, kan man beregne besparelser ved et (forventet) fald i adverse events og holde det op mod driftsomkostningerne ved pakkerne (i 'normal') drift.

**Table: Resultat-indikatorer for pakkerne**

<b>Pakke</b>	<b>Indikator</b>	<b>Senge-afdeling</b>	<b>Inten-siv</b>	<b>Kirur-gi</b>	<b>Medicin</b>
Mobilt akutsystem	Antal hjertestopkald per 1000 sengedage	x			
Mobilt akutsystem	Andel korrekt udførte basisobservationer	x			
Mobilt akutsystem	Procent observationsdøgn med afvigende basisobservationer og relevant reaktion på afvigelsen	x			
CVK-pakke	Dage mellem CVK-relaterede infektioner	x	x		
CVK-pakke	Procent af CVK'er som har opfyldt alle elementer for anlæggelse af CVK	x	x		
CVK-pakke	Procent CVK'er med relevant aktuel indikation	x	x		
Respiratorpakke	Dage mellem respiratorrelaterede lungeinfektioner		x		
Respiratorpakke	Procent hele respiratordøgn med opfyldelse af respiratorpakken		x		
Respiratorpakke	Procent reintubationer		x		
AMI-pakke	Procent AMI-patienter som dør under indlæggelse	x			
AMI-pakke	Procent AMI-patienter som modtager alle elementer i AMI-pakken	x			
Sepsisapakke	Antal patienter med svær sepsis eller septisk chok		x		
Sepsisapakke	Procent patienter med svær sepsis eller septisk chok, som har fået opfyldt alle elementer i 6-timerspakken		x		
Medicinafstemning	Procent indlæggelser hvor der er udført medicinafstemning	x			x
Medicinafstemning	Procent udskrivelser hvor der er udført medicinafstemning	x			x
Højriskomedicin-pak	Dage mellem alvorlige medicineringsfejl	x			x
Højriskomedicin-pak	Antal indlagte patienter med afvigende biokemiske markører	x			x
KAD-pakke	Dage mellem KAD-relateret urinvejsinfektion	x	x		
KAD-pakke	Antal KAD-døgn per 1000 sengedage	x	x		
KAD-pakke	Procent KAD med relevant indikation	x	x		
PVK-pakke	Procent perifere venekatetre som ved fjernelse viser tegn til inflammation eller er gået subcutant	x	x		
PVK-pakke	Procent perifere venekatetre med relevant og aktuel indikation	x	x		
PVK-pakke	Procent perifere venekatetre som fjernes før 72 timer	x	x		
Tryksårspakke	Procent patienter med sygehus erhvervede tryksår	x	x		
Tryksårspakke	Procent patienter som screenes for tryksårsrisiko ved indlæggelsen	x	x		
Tryksårspakke	Procent patienter med risiko for tryksår som revurderes for tryksår	x	x		
Hjertesvigtpakke	Procent patienter med nydiagnosticeret systolisk dysfunktion som genindlægges inden for 28 dage efter udskrivelse	x			
Hjertesvigtpakke	Procent patienter med nydiagnosticeret systolisk dysfunktion som modtager alle elementerne i hjertesvigtpakken	x			
Kirurgipakke	Procent opererede patienter som dør under indlæggelsen			x	
Kirurgipakke	Procent opererede patienter som genindlægges inden for 30 dage efter udskrivelse			x	
Kirurgipakke	Procent opererede patienter som modtager alle elementer i kirurgipakken			x	
Håndhygiejne	Procent situationer med krævet håndhygiejne, hvor håndhygiejne er korrekt udført	x	x		
Håndhygiejne	Forbrug af hånddesinfektionsprodukter	x	x		

## ***Litteratur***

- (1) Lynn J, Baily MA, Bottrell M, Jennings B, Levine RJ, Davidoff F, et al. The ethics of using quality improvement methods in health care. *Ann Intern Med* 2007 May 1;146(9):666-73.
- (2) Ting HH, Shojanian KG, Montori VM, Bradley EH. Quality improvement: science and action. *Circulation* 2009 Apr 14;119(14):1962-74.
- (3) Campbell D, Stanley J. *Experimental and Quasi-experimental Designs for Research*. Houghton Mifflin; 1966.
- (4) Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston New York: Houghton Mifflin Company; 2002.
- (5) Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000 Sep 16;321(7262):694-6.
- (6) Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. *BMJ* 2007 Mar 3;334(7591):455-9.
- (7) Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998;316:285.
- (8) Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in 6 therapeutic trials. *J Chronic Dis* 1967;20:637-48.:637-48.
- (9) Davidoff F, Batalden P. Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care* 2005 Oct;14(5):319-25
- (10) Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney SE. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *BMJ* 2009;338:a3152.
- (11) Berwick DM. Broadening the view of evidence-based medicine. *Qual Saf Health Care* 2005 Oct;14(5):315-6.
- (12) Solberg LI, Mosser G, McDonald S. The three faces of performance measurement: improvement, accountability, and research. *Jt Comm J Qual Improv* 1997 Mar;23(3):135-47.
- (13) Pronovost P, Wachter R. Proposed standards for quality improvement research and publication: one step forward and two steps back. *Qual Saf Health Care* 2006 Jun;15(3):152-3.
- (14) Davies P, Walker AE, Grimshaw JM. A systematic review of the use of theory in the design of guideline dissemination and implementation strategies and interpretation of the results of rigorous evaluations. *Implement Sci* 2010;5:14.
- (15) Francis JJ, Tinmouth A, Stanworth SJ, Grimshaw JM, Johnston M, Hyde C, et al. Using theories of behaviour to understand transfusion prescribing in three clinical contexts in two countries: Development work for an implementation trial. *Implement Sci* 2009;4:70.

- (16) Francis JJ, Stockton C, Eccles MP, Johnston M, Cuthbertson BH, Grimshaw JM, et al. Evidence-based selection of theories for designing behaviour change interventions: using methods based on theoretical construct domains to understand clinicians' blood transfusion behaviour. *Br J Health Psychol* 2009 Nov;14(Pt 4):625-46.
- (17) Li LC, Moja L, Romero A, Sayre EC, Grimshaw JM. Nonrandomized quality improvement intervention trials might overstate the strength of causal inference of their findings. *J Clin Epidemiol* 2009 Sep;62(9):959-66.
- (18) Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff (Millwood)* 2005 Jan;24(1):138-50.
- (19) Dansk Selskab for Patientsikkerhed, Anhøj J. Målsætninger for Patientsikkert Sygehus (måle-strategi). Notat 2010.
- (20) Black N. Assessing the quality of hospitals. *BMJ* 2010;340:c2066.
- (21) Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010;340:c2016.
- (22) Naessens JM, Campbell CR, Huddleston JM, Berg BP, Lefante JJ, Williams AR, et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009 Aug;21(4):301-7
- (23) Rabøl LI, Jølving LR, Bjørn, Viskum B. Systematisk journalgennemgang med IHI's GLOBAL TRIGGER TOOL. Erfaringer og resultater fra afprøvning på danske sygehuse. Middelfart: Center for Kvalitet; 2008.
- (24) Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Health Care* 2003 Dec;12 Suppl 2:ii39-ii45.
- (25) Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. *Jt Comm J Qual Patient Saf* 2006 Oct;32(10):585-90.
- (26) U.S.Department of Health and Human Services Agency for Healthcare Research and Quality. Triggers and Targeted Injury Detection Systems (TIDS) Expert Panel Meeting: Conference Summary . Rockville: U. S. Department of Health and Human Services Agency for Healthcare Research and Quality ; 2009.