

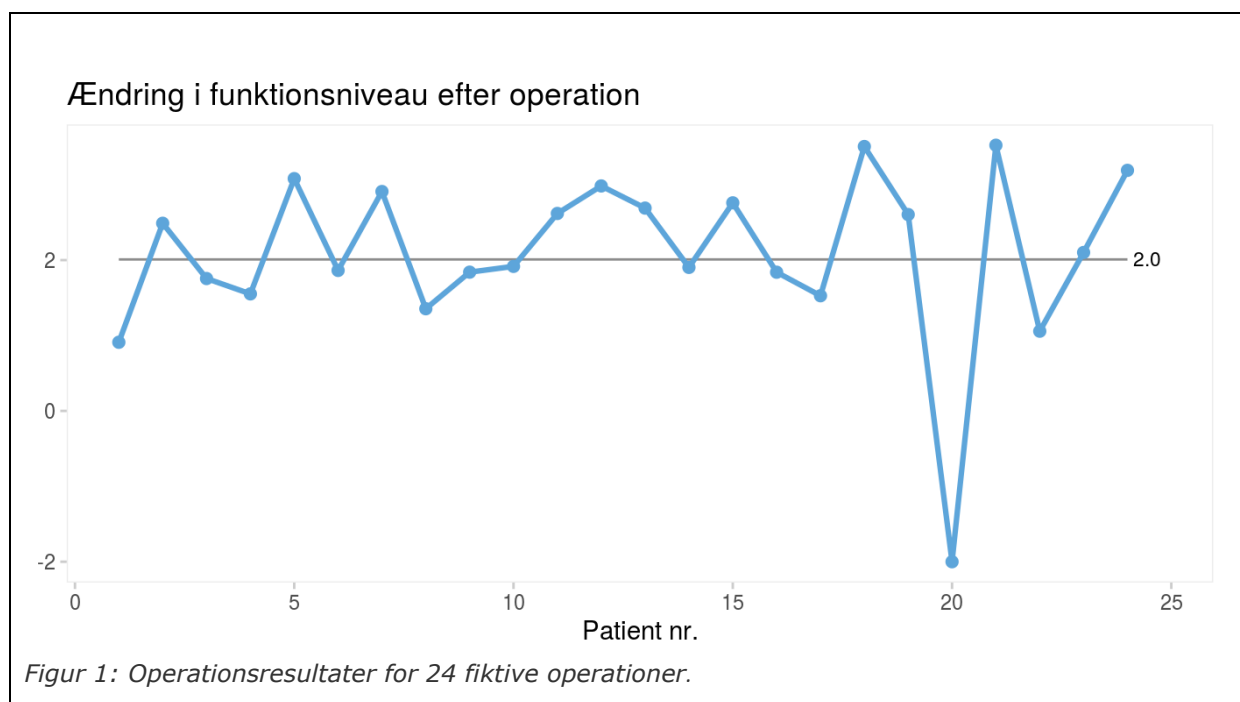
SPC for læger

Af Jacob Anhøj, overlæge, Diagnostisk Center, Rigshospitalet

Statistisk proces kontrol, SPC, er en gren af den analytiske statistiske videnskab, som bruges inden for kvalitetsudvikling til at forbedre produkter og tjenesteydelser. Centralt i SPC er forståelse for processers variation over tid og sted. Procesbegrebet skal i denne sammenhæng forstås bredt som noget, der modtager input og leverer et resultat. En proces kan fx være en operation og resultatet et udtryk for i hvor høj grad operationen fx afhjælper patienternes smerter. Af hensyn til den statistiske analyse er det nødvendigt, at resultatet kan udtrykkes i tal. Tal, som på denne måde udtrykker kvalitet, kalder vi for **indikatorer**.

Som et tænkt eksempel kunne vi lade nogle patienter score deres eget behandlingsresultat, fx funktionsniveau, på en visuel analogskala fra -5 til 5, hvor værdier over 0 udtrykker en forbedring, og værdier under 0 udtrykker en forværring.

På grund af variation i utallige faktorer, som har betydning for resultatet, er indikatorværdien sjældent helt ens for to patienter. I nogle tilfælde kan forskelle i behandlingsresultater forklares ved særlige forhold, fx alternativ operationsteknik, konkurrerende sygdomme hos patienten eller defekt udstyr, som bevirkede, at det gik særligt godt eller skidt for en patient. Men i de fleste tilfælde er det ikke muligt entydigt at forklare de forskelle, vi observerer, som andet end naturlig (også kaldet 'tilfældig') variation.



Diagrammet i figur 1 viser operationsresultatet fra 24 fiktive operationer. Prikkerne på kurven er afsat i den rækkefølge, patienterne blev opereret. Den vandrette streg markerer den mediane funktionsforbedring, som er 2. Selv om resultatet varierer fra patient til patient, ligger de fleste i nærheden af midtpunktet. Men patient nr. 20 skiller sig ud med en score på -2. En samvittighedsfuld kirurg vil straks spørge "hvorfor?". Der kan være

mange forklaringer, og langt fra alle handler om dårlig operationsteknik. Det væsentlige er at finde årsagen og så vidt muligt forhindre, at det sker igen.

Kvalitetsindikatorer knyttes ofte til kvalitets**standarder**, som udtrykker det ønskede eller forventede niveau for kvaliteten. Standarder kan være enkeltværdier, som udtrykker det lavest eller højest acceptable niveau, eller dobbeltværdier, som udtrykker et ønsket interval for en given indikator (fx INR i et terapeutisk interval).

Det er en vigtig overvejelse, som ofte glemmes, om en standard gælder enkeltmålinger eller aggregerede målinger. Hvis standarden gælder enkeltmålinger, tilstræber vi, at alle målinger ligger på den rigtige side af standarden. Hvis standarden gælder aggregerede målinger, er det tilstrækkeligt, at gennemsnittet eller medianen af alle målingerne ligger rigtigt. Hvis fx standarden for operationsresultatet fastsættes til 1,5, skal vi vide, om den kun er opfyldt, hvis alle patienter scorer over 1,5, eller om det er tilstrækkeligt, at den gennemsnitlige forbedring er over 1,5.

Det er også værd at overveje, om man skal stille krav til variationsbredden – altså, hvor meget må enkeltmålingerne afvige fra hinanden og fra gennemsnittet? De fleste vil nok foretrække så lidt variation som muligt så fx en Big Mac eller en cola smager ens i Hangzhou og i Herning.

God kvalitet betyder derfor to ting: 1: at resultatet varierer mindst muligt og på en forudsigelig måde, og 2: at resultatet lever op til de forventninger, der knyttes til det.

To slags variation

Hovedformålet med at analysere data over tid er at erkende væsentlige ændringer i de underliggende strukturer og processer, som skaber data, så vi kan handle fornuftigt, hvis ændringerne tyder på en forringelse af kvaliteten af vores ydelse.

Men selv om tallene ændrer sig, betyder det ikke nødvendigvis, at den underliggende proces har ændret sig. Og selv når processen ændrer sig, er det ikke altid, at tallene ændrer sig. I eksemplet ovenfor vil de færreste være i tvivl om, at patient nr. 20 repræsenterer en afvigelse. Men ville vi også opfatte det som en afvigelse, hvis patient 20 havde scoret 0? Så hvordan skelner vi mellem tilfældige måleudsving og sikre forandringer?

Walther A. Shewhart, som grundlagde SPC, beskrev to slags variation: almindelig variation (*common cause variation*) og særlig variation (*special cause variation*) (Shewhart 1931).

Almindelig variation

- kaldes også tilfældig variation eller støj,
- findes i alle processer,
- skyldes fænomener, som altid er til stede,
- identificerer en stabil og dermed forudsigelig proces.

Særlig variation

- kaldes også ikke-tilfældig variation eller signal,
- findes i nogle men ikke alle processer,
- skyldes fænomener, som ikke normalt er til stede i systemet,
- identificerer en ustabil og dermed uforudsigelig proces.

Evnen til at skelne almindelig og særlig variation fra hinanden er afgørende for valget af forbedringsstrategi. Er processen ustabil (= særlig variation), bør man søge at finde årsagen eller årsagerne til ustabiliteten og eliminere eller implementere disse alt efter, om afvigelserne går den forkerte eller rigtige vej. Målet er at stabilisere processen på det ønskede niveau.

Er processen derimod stabil (= almindelig variation), skal man spørge sig selv, om den er tilfredsstillende: Opfylder den "kundens" (patientens, ledelsens, borgerens, samfundets osv.) behov og forventninger? Hvis det er tilfældet, er det vigtigt at beholde hænderne i lommen og undgå at forstyrre processen, selv om målingerne *tilfældigt* går op og ned. At reagere på tilfældige udsving kaldes *tampering* og kan medføre øget variation (Quality America Inc. 2016).

Er processen stabil men utilfredsstillende, er det nødvendigt med gennemgribende forandringer i de underliggende strukturer og arbejdsgange, som driver processen, for at flytte den til det ønskede niveau.

	Stabil proces	Ustabil proces
Tilfredsstillende resultat	<p>Alt i orden:</p> <p><i>Kontrollér processen: Overvåg processen, minimér uønsket, tilfældig variation.</i></p>	<p>Procesproblemer:</p> <p><i>Stabilisér processen: Eliminér uønsket eller implementér ønsket ikke-tilfældig variation.</i></p>
Utilfredsstillende resultat	<p>Produktproblemer:</p> <p><i>Revidér processen: Optimér strukturer, procedurer og arbejdsgange, så processen flytter sig og stabiliseres på et nyt og bedre niveau.</i></p>	<p>Kaos:</p> <p><i>Stabilisér processen: Eliminér uønsket eller implementér ønsket ikke-tilfældig variation.</i></p> <p>og</p> <p><i>Revidér processen: Optimér strukturer, procedurer og arbejdsgange, så processen flytter sig og stabiliseres på et nyt og bedre niveau.</i></p>

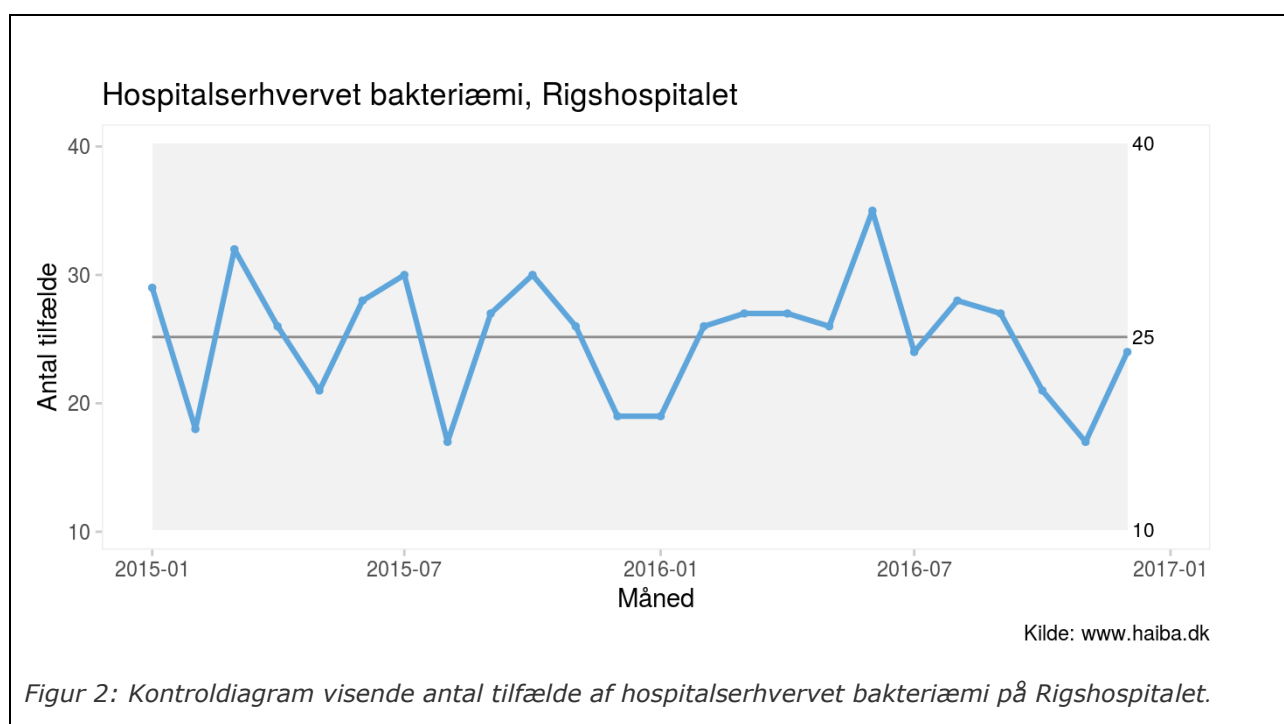
Tabel 1: Tabellen viser de fire tilstande, som en proces kan befinde sig i, og de tilhørende passende handlinger.

Shewharts kontrolidiagram

Shewhart opfandt kontrolidiagrammet i 1924 som et enkelt redskab til at skelne mellem almindelig og særlig variation i produktionsprocesser. Shewhart arbejdede på det tidspunkt i inspektionsafdelingen på Bell Labs, som producerede telefonapparater. Han beskrev i sin første bog fra 1931 en lineær og ineffektiv metode til kvalitetskontrol: specifikation → produktion → inspektion. Man inspicerede simpelthen alle færdige

produkter og kasserede eller reparerede dem, der ikke overholdt specifikationerne. Shewhart fandt ud af, at man ved at indarbejde kvalitetsstyring i alle led af produktionskæden, kunne minimere eller ligefrem undgå, defekte produkter i sidste ende og dermed spare virksomheden for mange resurser. Det var til dette formål, han opfandt kontrolldiagrammet, som gjorde det muligt for medarbejderne ved samlebåndet at vide, hvornår og hvordan (som i tabel 1) de skulle reagere på afvigelser i løbende kvalitetsmålinger i de forskellige led af produktionskæden.

Kontrolldiagrammet er et xy-diagram med prikker forbundet med lige streger. X-aksen viser tiden eller rækkefølgen af målingerne, og y-aksen viser indikatorværdierne. En vandret linje markerer datas midtpunkt (som regel gennemsnittet), og grænserne for den almindelige variation vises som en øvre og nedre kontrolgrænse markeret med grå baggrund i figuren.



Kontrolgrænserne placeres normalt ved ± 3 standardafvigelser (SD eller 'sigma') fra midtlinjen. I figuren ovenfor er øvre og nedre kontrolgrænse hhv. 10 og 40. Bemærk, at standardafvigelsen, som indgår i beregningen, er den estimerede almindelige variation – ikke den puljede standardafvigelse af alle datapunkterne, som også omfatter eventuel særlig variation. For at undgå misforståelser bruger de fleste betegnelsen sigma om den standardafvigelse, som indgår i beregning af kontrolgrænser.

Udregning af kontrolgrænser forudsætter antagelser om datas teoretiske sandsynlighedsfordeling, og der findes mange forskellige typer kontrolldiagrammer til forskellige typer data. Men fortolkningen er enkel og ens for dem alle: Hvis alle målepunkterne ligger mellem kontrolgrænserne, er der almindelig variation. Målepunkter uden for kontrolgrænserne indikerer særlig variation. Vi vender tilbage til beregning af kontrolgrænser i afsnittet om valg af kontrolldiagram.

Kontrolldiagrammet ovenfor viser den månedlige forekomst af tilfælde af hospitalserhvervet bakteriæmi på Rigshospitalet i perioden 2015-2016. I gennemsnit er

der 25 tilfælde om måneden. De fleste måneder er der mellem 20 og 30 tilfælde og enkelte måneder lidt flere eller færre. Kontrolgrænserne viser, at den almindelige variation spænder fra 10 til 40 tilfælde om måneden. Med andre ord: hvis intet forandrer sig, skal vi i fremtiden ikke undre os, hvis der en enkelt måned er så få som 10 eller så mange som 40 tilfælde.

Det er vigtigt at forstå, at kontrolgrænserne beregnes ud fra den naturlige variation, dvs. de fortæller, hvad processen er i stand til at levere, hvad enten vi bryder os om det eller ej. Kontrolgrænser er altså noget ganske andet end specifikationsgrænser og standarder, som taler med kundens stemme dvs., hvilket niveau vi ønsker os. Nogen vil måske mene, at grænserne for den naturlige variation i bakteriemital er uacceptabelt brede. Men det er der ikke noget at gøre ved, sådan fungerer processen, og hvis man ønsker andre grænser, må man ændre processen.

God kvalitet – som i øverste venstre kvadrant i tabel 1 – er når processen og kunden taler med én stemme, dvs. når processen er stabil, og standarden (dvs. målsætningen) ligger på den rigtige side af midtlinjen eller kontrolgrænserne afhængig af, om den vedrører enkeltmålinger eller aggregerede målinger. Hvis fx Rigshospitalet beslutter, at mere end 35 hospitalserhvervede bakteriemier på en måned er uacceptabelt, er målet at bringe øvre kontrolgrænse ned under 35 ved at rykke hele processen ned. Hvis standarden derimod gælder gennemsnitsniveauet over længere tid, er alt i orden, fordi gennemsnittet på 25 ligger bekvemt under standarden.

Statistiske test til påvisning af særlig variation

Almindelig variation er karakteriseret ved, at selv om vi ikke kan forudsige den præcise placering af det næste datapunkt, fordeler datapunkterne sig forudsigeligt omkring et givet centrum og inden for visse grænser. Særlig variation betyder, at processen ændrer sig, så vi ikke længere kan forudsige centrum og/eller grænser (før en ny, stabil proces er etableret).

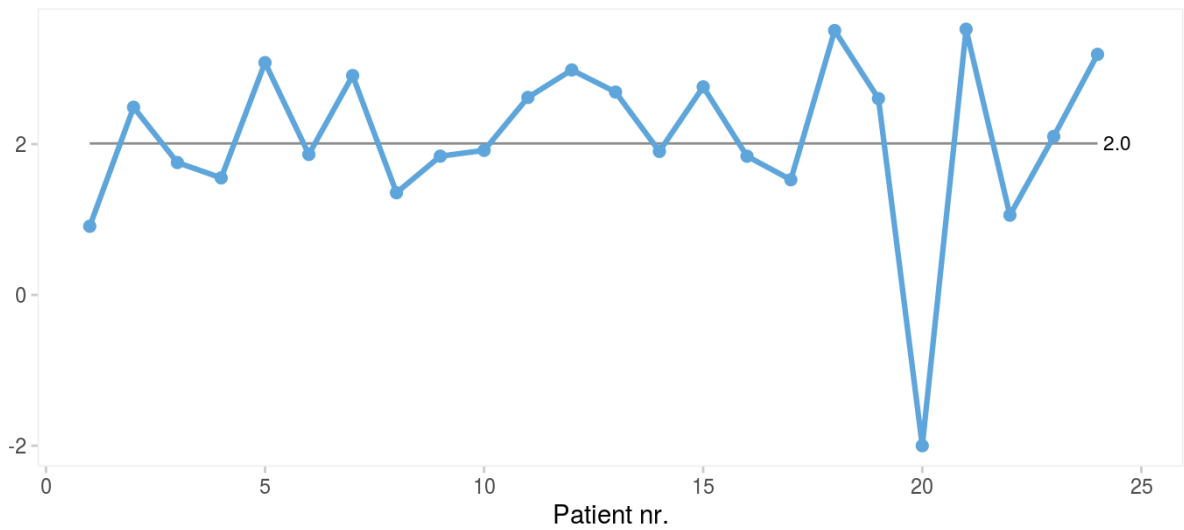
Tag fx en almindelig sekssidet terning. Det er umuligt at forudsige resultatet af det næste kast; men vi ved, at fremtidige kast vil antage værdier mellem 1 og 6, at sandsynligheden for hvert muligt udfald er $1/6$, og at gennemsnittet ligger tæt på 3,5. Hvis en terning pludselig viser 7, vil vi undre os. Eller tilsvarende, hvis terningen pludselig viser samme værdi mange gange i træk, ved vi, at der er sket noget usædvanligt, og at det vil kunne betale sig at lede efter årsagen.

Test for særlig variation bygger på påvisning af mønstre i data, som ville være usædvanlige i tilfældige og forudsigelige processer.

Shewharts 3 sigma-regel

Shewharts oprindelige 3 sigma-regel (et eller flere datapunkter ligger uden for kontrolgrænserne) er effektiv til at opdage større, eventuelt forbigående udsving i processen. Patient nr. 20 fra kirurgieksemplet (figur 3) falder tydeligt uden for nedre 3 sigma-kontrolgrænse og repræsenterer derfor særlig variation.

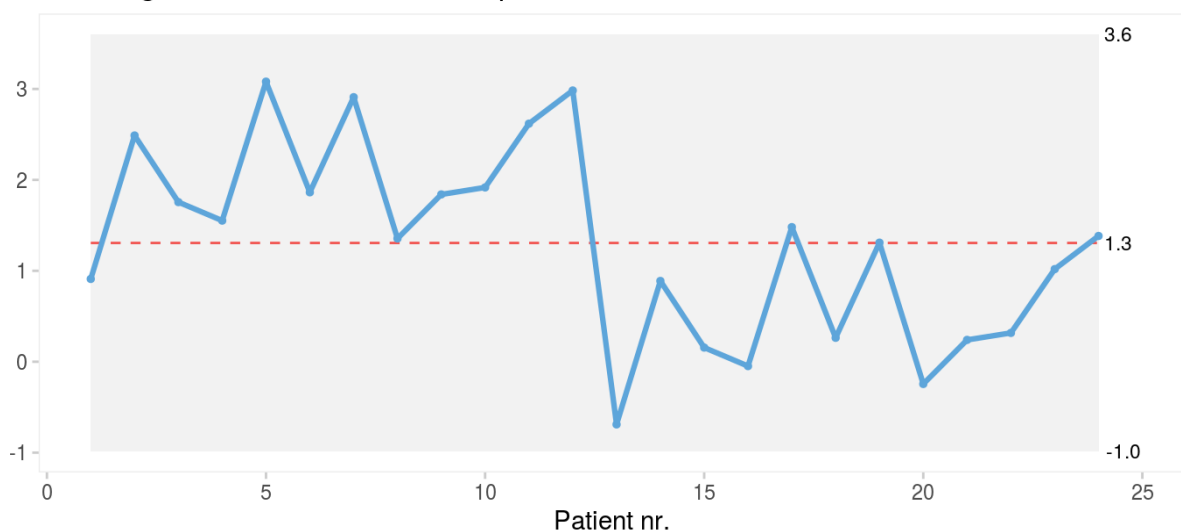
Ændring i funktionsniveau efter operation



Figur 1: Operationsresultater for 24 fiktive operationer

3 sigma-reglen er derimod ikke effektiv til at opdage mindre skift (dvs. ændringer) i processen som i næste tænkte eksempel (figur 4), hvor der tydeligvis sker et skift efter patient nr. 12, selv om alle datapunkterne ligger mellem kontrolgrænserne.

Ændring i funktionsniveau efter operation



Figur 4: Operationsresultater for 24 fiktive operationer (ny patientserie).

Derfor er der i tidens løb udviklet mange supplerende test til at øge kontroldiagrammers mulighed for at afsløre andre former for skift i kvaliteten end pludselige store udsving.

Western Electric-reglerne

De bedst kendte test for særlig variation er formentlig *The Western Electric Rules* (WE),

som optræder første gang i *Statistical Quality Control Handbook* i 1956 (Western Electric Company 1956). WE-reglerne er fire enkle tests, som kan appliceres på kontroldiagrammet alene ved at kigge på det og baserer sig på at finde usædvanlige mønstre i datapunkternes placering i forhold til kontrolgrænserne og midtlinjen.

1. Et eller flere datapunkter ligger uden for kontrolgrænserne (Shewharts oprindelige 3 sigma-regel).
2. To ud af tre konsekutive datapunkter er mere end 2 sigma fra midtlinjen (to tredjedele af afstanden mellem midtlinje og kontrolgrænsen).
3. Fire ud af fem konsekutive datapunkter er mere end 1 sigma fra midtlinjen.
4. Otte konsekutive datapunkter ligger på samme side af midtlinjen.

Figur 4 er således positiv for både regel 2, 3 og 4.

WE-reglerne øger kontroldiagrammets følsomhed for mindre, men vedvarende skift i data, og dermed kan det anvendes til at identificere mindre men mere vedvarende ændringer i fx kvaliteten af en ydelse. Reglerne har vist deres værd gennem mere end et halvt århundrede og er standard i mange virksomheder verden over.

Det er værd at bemærke, at WE-reglerne fungerer bedst, når kontroldiagrammet har mellem 20 og 30 datapunkter. Med færre end 20 datapunkter, mister reglerne sensitivitet (flere falsk negative), og med flere end 30 datapunkter mister de specificitet (flere falsk positive).

I Western Electric-håndbogen findes i øvrigt mange flere regler til påvisning af særlige typer af særlig variation. Men det er vigtigt at være opmærksom på, at selv om det kan være fristende at bruge mange tests for at øge følsomheden af sin analyse, stiger risikoen for falske signaler med antallet af test – jo flere regler, man applicerer, jo flere falske signaler skal man være villig til at forholde sig til. Det behøver ikke være et problem, hvis omkostningerne ved årsagsanalyse er beskedne. Men hvis et signal udløser fx produktionsstop eller dyre kerneårsagsanalyser, skal man nøje opveje omkostningerne ved at overse et skift mod omkostningerne ved falske signaler. Beslutningen om, hvor mange og hvilke test, der skal indgå i analysen er derfor vigtig og skal baseres på meget andet end statistiske overvejelser.

Test for lange serier og få kryds

To nyere regler for særlig variation baserer sig alene på datapunkternes placering i forhold til midtlinjen (Anhøj og Olesen 2014; Anhøj 2015):

- **Skiftreglen: Usædvanligt lange serier.** En serie (*run*) er et eller flere konsekutive datapunkter på samme side af midtlinjen. Datapunkter, som ligger direkte på midtlinjen hverken bryder eller bidrager til serien. Den øvre 95% prædiktionsgrænse for længste serie er omtrent $\log_2(n)+3$ afrundet til nærmeste heltal, hvor n er antallet af brugbare datapunkter (datapunkter som ikke ligger på midtlinjen). Med 24 datapunkter vil en serie på mere end 8 datapunkter indikere særlig variation.
- **Krydsreglen: Usædvanligt få kryds.** Et kryds er når to nabopunkter ligger på hver sin side af midtlinjen. I en tilfældig proces vil antallet af kryds være binomialfordelt, $b(n-1, 0.5)$, hvor n er antallet af brugbare observationer og 0.5 er sandsynligheden for, at et datapunkt befinder sig hhv. over eller under midtlinjen. Den nedre 5% prædiktionsgrænse for antal kryds er den kumulative

fordelingsfunksjons nedre 5% grænse. Med 24 datapunkter vil færre end 8 kryds indikere særlig variation.

[Grænseværdier](#) for længste serie og antal kryds i tilfældige talrækker med 10-100 datapunkter.

Skift- og krydsreglerne er bedre end 3 sigma-reglen og sammenlignelige med WE-regel 2-4 til at påvise moderate men *vedvarende* skift i data over tid (Anhøj og Wentzel-Larsen 2018). Derudover har de nogle ekstra fordele:

- De er uafhængige af kontrolgrænser og kan derfor bruges alene i såkaldte seriediagrammer, som er enklere at konstruere og ikke forudsætter antagelser om datas teoretiske sandsynlighedsfordeling. Vi ser nærmere på seriediagrammet nedenfor.
- Fordi grænseværdierne for længste serie og antal kryds af midtlinjen afhænger af antallet af datapunkter, holder reglerne nogenlunde konstant sensitivitet og specifitet i diagrammer med så få som 10 og op til uendeligt mange datapunkter.

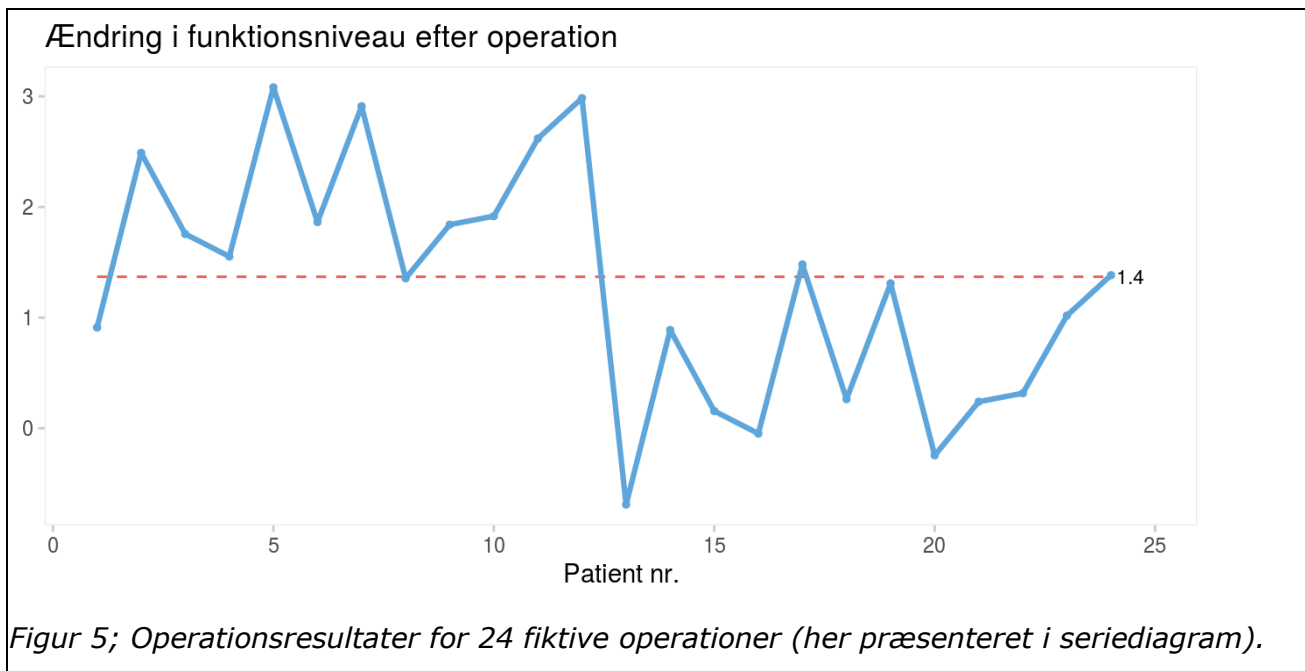
Serielængde og antal kryds er to sider af samme sag – lange serier giver færre kryds og omvendt – og begge regler, sammen eller hver for sig, indikerer særlig variation. Reglerne kan således opfattes som alternativer til WE-regel 2-4. De kan benyttes sammen med 3 sigma-reglen i kontrolprogrammer eller alene i seriediagrammer.

I det foregående kontrolprogram (figur 4), hvor alle datapunkterne ligger mellem kontrolgrænserne, finder vi særlig variation i form af en lang serie på 11 datapunkter (mod forventet højst 8) og kun 7 kryds (mod forventet mindst 8). Bemærk den røde, stiplede midtlinje, som indikerer ikke-tilfældig variation i form af enten for lange serier eller for få kryds.

Seriediagrammet

Skift- og krydsreglerne kan som sagt bruges alene i seriediagrammer, som ligner kontrolprogrammer, men mangler kontrolgrænser, og hvor midtlinjen som regel er medianen (dvs. Er placeret som en vandret linje svarende til den midterste værdi af datapunkterne). Med seriediagrammer behøver man derfor ikke besvære sig med komplicerede udregninger af kontrolgrænser; og med medianen som centrum behøver man ikke – som med kontrolprogrammer – bekymre sig om datas fordeling, som per definition altid vil være symmetrisk omkring medianen.

Figur 5 viser et seriediagram med samme datasæt som i kontrolprogrammet i figur 4.



Seriediagrammet er i sagens natur uegnet til at opdage selv store, men forbigående skift, som ellers ville blive fanget af kontrolgrænser. Men på grund af sin enkelthed og mere sikre påvisning af vedvarende skift, er seriediagrammet velegnet som førstevalg. Kontrolldiagrammet kan så anvendes i de tilfælde, hvor seriediagrammet viser almindelig variation, og hvor det er vigtigt at udelukke større forbigående skift i processen. Hvis seriediagrammet finder særlig variation, er der ingen grund til at beregne kontrolgrænser. Ligesom der sjældent er grund til CT-scanning, hvis diagnosen kan stilles sikkert med en almindelig røntgenundersøgelse.

Serie- og/eller kontrolldiagrammer – hvad skal du vælge?

I kvalitetsudviklingslitteraturen får man let det indtryk, at kontrolldiagrammer er finere eller bedre end seriediagrammer, hvis eneste fordel er, at de er enkle at konstruere. Det passer bare ikke. Serie- og kontrolldiagrammer er forskellige diagnostiske metoder, som komplementerer hinanden. Derfor benytter vi fællesbetegnelsen SPC-diagrammer. Beslutning om, hvorvidt midtlinjen skal være medianen (som i seriediagrammet) eller gennemsnittet (som i kontrolldiagrammet), om man skal bruge kontrolgrænser eller ej, og hvilke test der skal udføres, afhænger ikke af hvilket diagram, der er lettest at konstruere, men af analysens formål.

Til kvalitetsudvikling, hvor målet er at flytte en proces fra et niveau til et andet og bedre niveau (og ikke bare at overvåge kvaliteten), og hvor vi derfor forventer særlig variation, kan man med fordel applicere skift og kryds-reglerne i forhold til medianen. Hvis skift og kryds-reglerne signalerer, kan vi med stor sikkerhed sige, at processen har flyttet sig mere end bare forbigående.

Til kvalitetskontrol, når målet er at fastholde en tilfredsstillende proces på et godt og stabilt niveau, og hvor vi forventer almindelig variation, kan man med fordel benytte gennemsnittet som midtlinje og supplere med kontrolgrænser, som er hurtigere til at opdage større udsving i data.

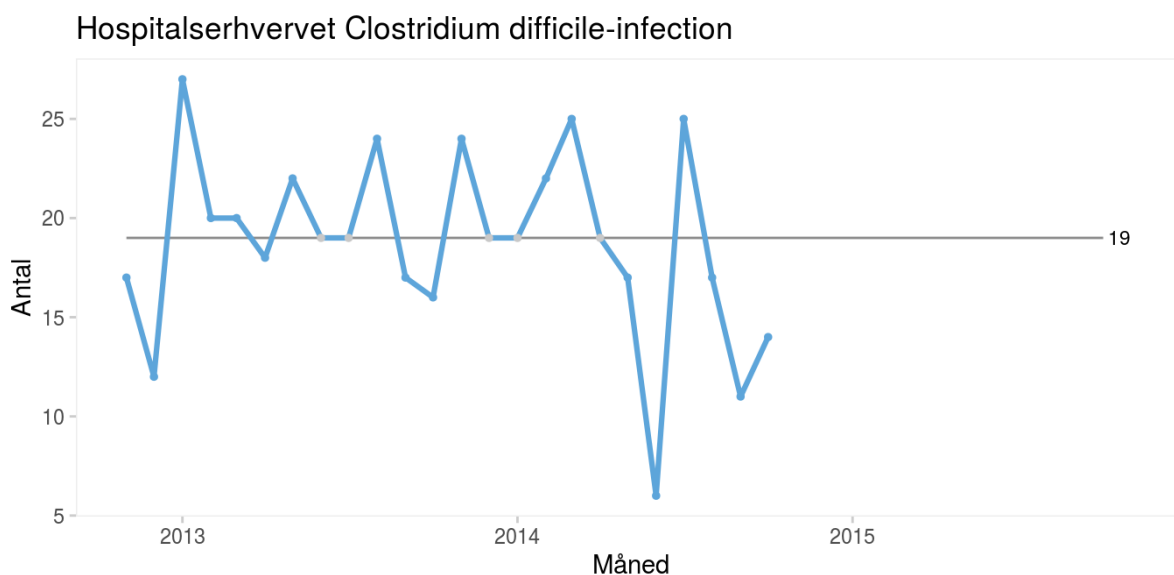
I praksis kan man vælge følgende strategi:

1. Beslut målsætning og vælg indikatorer.
2. Definer indikatorerne (mere om dette i afsnittet om indikatorer).
3. Indsamle data og begynd seriediagram. Dette kan gøres enkelte fx ved hjælp af papir og blyant.
4. Efter mindst 12 datapunkter, test for særlig variation i form af usædvanligt lange serier eller usædvanligt få kryds:
5. almindelig variation: fastholde eller forbedre?
6. særlig variation: eliminere eller implementere?
7. Forbedring er opnået og fastholdt, når de seneste mindst 12 målepunkter varierer tilfældigt omkring et nyt og bedre centrum.
8. Overvej at anvende et kontrol-diagram til at overvåge, stabilisere og fastholde processen.

Case: Forebyggelse af Clostridium difficile-infektioner hos indlagte patienter

På Amager-Hvidovre Hospital besluttede man i efteråret 2014 på udvalgte afdelinger at afprøve profylaktisk anvendelse af gærsvampen *Saccharomyces boulardii* (fås i kapselform i helsekostbutikker) mod *Clostridium difficile*-infektion hos patienter i antibiotisk behandling (Carstensen m.fl. 2018). Interventionen er velbeskrevet, så formålet var ikke at bevise effekten, men at implementere en ny praksis og reducere risikoen for hospitalserhvervet *C. difficile*-infektion.

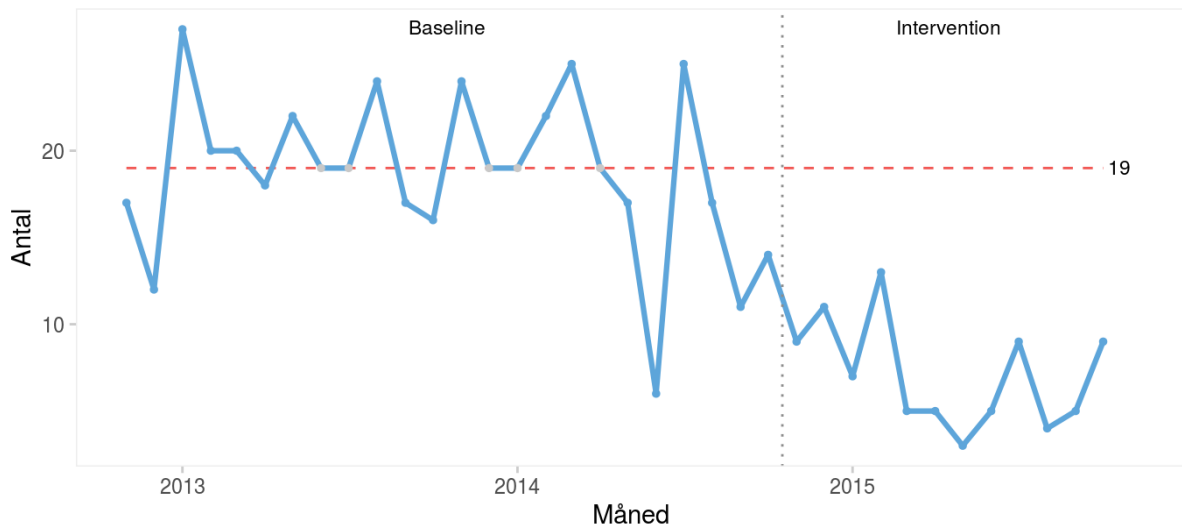
Seriediagrammet viser det månedlige antal *C. difficile*-infektioner på hele hospitalet 24 måneder inden iærksættelse af interventionen (baseline).



Fordi der er tilfældig variation, kan vi låse medianen og forlænge den ud i fremtiden. Medianen repræsenterer det forventede fremtidige niveau, hvis intet ændrer sig. Der er 19 brugbare observationer, længste serie er på 3 datapunkter (øvre grænse = 7), og kurven krydser medianen 8 gange (nedre grænse = 6). Diagrammet viser tilfældig variation omkring 19 tilfælde om måneden.

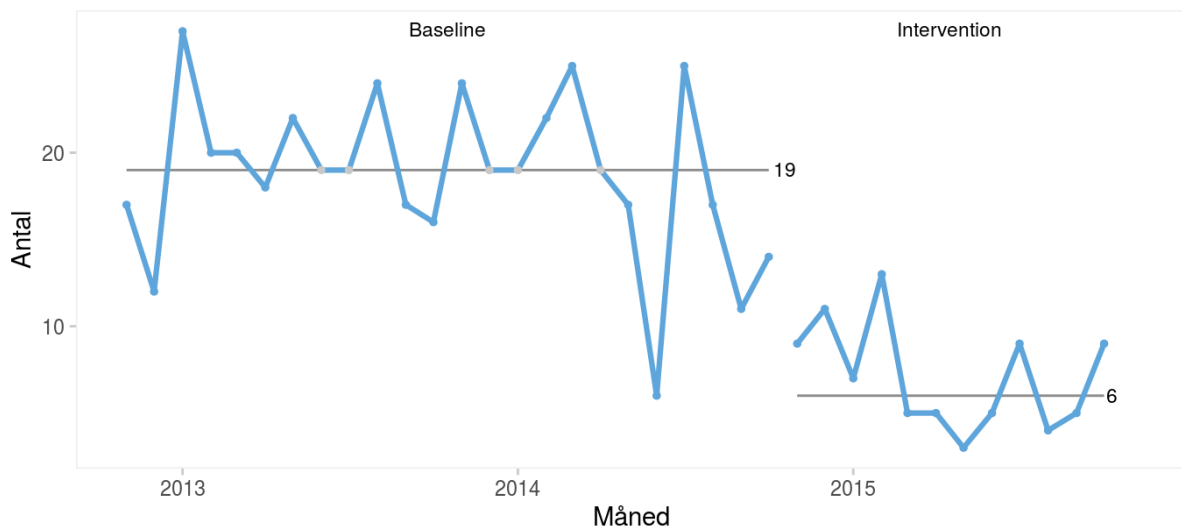
Interventionen begyndte i oktober 2014, og det er tydeligt, at der efterfølgende sker et skift i processen mod færre infektioner. Skiftet signaleres af både en usædvanlig lang serie ($15 > 8$) og usædvanligt få kryds ($8 < 11$).

Hospitalserhvervet Clostridium difficile-infection



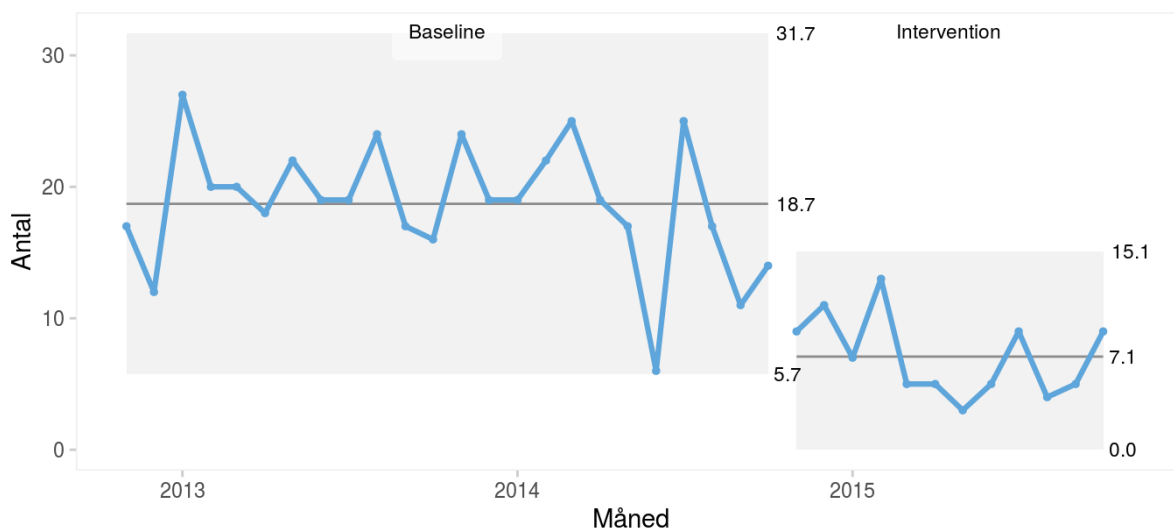
Når processen skifter i den ønskede retning som resultat af en bevidst indsats, kan det være meningsfuldt at opdele diagrammet i to perioder før og efter interventionen.

Hospitalserhvervet Clostridium difficile-infection



Fordi begge perioder kun indeholder almindelig variation, kan det være nyttigt at benytte et kontrol-diagram til at vise den nye proces' naturlige grænser.

Hospitalserhvervet Clostridium difficile-infection



Kontroldiagrammet viser, at infektionsraten er faldet fra i gennemsnit 19 til 7 tilfælde om måneden efter interventionen, at den nye proces er stabil og dermed forudsigelig, og at vi i fremtiden kan forvente mellem 0 og 15 tilfælde om måneden.

Selv om det er meget fristende at konkludere, at interventionen har medført en forbedring, kan man ikke sige det med sikkerhed. Alt, vi ved, er, at forekomsten af *C. difficile*-infektioner er faldet, efter at vi ændrede praksis. Hvis vi ønsker at bevise en kausal sammenhæng mellem en intervention og et resultat, må vi på anden vis føre bevis for sammenhængen, fx ved at lave et godt gammeldags kontrolleret, randomiseret forsøg eller ved at inddrage Bradford Hills kriterier i argumentationen (se kapitel 14 om publicering).

Formålet med løbende kvalitetsudvikling er ikke at bevise årsagssammenhænge men at forbedre kvaliteten. Det gør vi i dette tilfælde ved at implementere kendt viden i form af *S. boulardii* til forebyggelse af *C. difficile*-infektioner.

Valg af kontroldiagram

Der er udviklet mange forskellige typer kontroldiagrammer til forskellige datatyper. Forskellen ligger i måden, kontrolgrænserne skal beregnes på og afhænger bl.a. af, om data er måle- eller tælledata (se afsnit om måle- og tælledata). Tabellen viser de mest benyttede kontroldiagrammer til gængse datatyper; men der er mange flere til særlige formål, og det anbefales, at man konsulterer speciallitteraturen (fx Montgomery 2009), hvis man for alvor skal have gavn af kontroldiagrammer. I næste afsnit om indikatorer findes en nærmere beskrivelse af datatyper.

Den generelle formel til beregning af kontrolgrænser er $\bar{x} \pm 3SD$, hvor \bar{x} er gennemsnittet af alle datapunkterne, og SD er den estimerede standardafvigelse for den almindelige variation.

Valg af kontrolldiagram: \bar{x} = gennemsnit, $\bar{\bar{x}}$ = vægtet gennemsnit, \overline{MR} = gennemsnitlige, absolutte forskel mellem nabomålinger, A_3 = konstant som afhænger af stikprøvestørrelsen, \bar{s} = vægtet gennemsnitlige standardafvigelse, \bar{c} = gennemsnitlige antal defekter, n_i = udfaldsrummets/stikprøvens størrelse, \bar{p} = gennemsnitlige andel defekte enheder.

Diagram	Data	Antaget sandsynlighedsfordeling	Kontrolgrænser
<i>Måledata</i>			
I	Individuelle målinger	Gaussisk	$\bar{x} \pm 2.66\overline{MR}$
Xbar	Gennemsnit af flere målinger	Gaussisk	$\bar{\bar{x}} \pm A_3\bar{s}$
<i>Tælledata</i>			
C	Antal defekter	Poisson	$\bar{c} \pm 3\sqrt{\bar{c}}$
U	Defektrater	Poisson	$\bar{u} \pm 3\sqrt{\frac{\bar{u}}{n_i}}$
P	Andel defekte enheder	Binomial	$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$
G	Enheder mellem defekte enheder	Geometrisk	$\bar{x} \pm 3\sqrt{\bar{x}(\bar{x} + 1)}$

Tabel 2: viser karakteristika for forskellige typer af diagrammer

Begreberne defekte og defekter er statistiker-jargon for det, man tæller. Man taler om defekte enheder, når man tæller genstande, individer, processer osv., som besidder eller ikke besidder en bestemt egenskab, der som oftest – men ikke nødvendigvis – er uønsket.

En defekt enhed kan fx være en patient med tryksår (et eller flere), en operation som gik galt eller et lokale, som er utilfredsstillende rengjort. Defekte er altså et enten-eller-begreb og opfører sig i naturen ofte tilnærmelsesvist binomialfordelt.

Når man tæller defekter, tæller man fænomener, som er tilfældigt fordelt i tid og rum. En defekt kan fx være et tryksår, en hospitalsinfektion, en plet på gulvet eller et stjernesku. Defekter opfører sig ofte tilnærmelsesvist poissonfordelt.

Det er vigtigt at vælge det rigtige kontrolldiagram til sine data. SPC-software er ukritisk og vil fremstille det diagram, man beder om, uanset om det er passende eller ej. I nogle tilfælde får man nogenlunde samme resultat med forskellige kontrolldiagrammer; men ofte får man misvisende kontrolgrænser, hvis ikke diagramtypen passer til data.

Er man i tvivl om valget af kontrolldiagram, eller skal man bare have et par hurtige, "håndlavede" kontrolgrænser på sit seriediagram, kan man (næsten) altid benytte I-diagrammet, som er kontrolldiagrammernes "schweizerkniv". I-diagrammet baserer sine kontrolgrænser på den faktiske variation mellem nabomålepunkter og kan i praksis benyttes til de fleste typer data, hvis blot udfaldsrummet (nævneren) er nogenlunde konstant. Først beregner man gennemsnittet (= midtlinjen) af de individuelle datapunkter. Dernæst ganger man den gennemsnitlige absolutte forskel mellem nabomålinger med den statistiske konstant 2,66. Så beregner man kontrolgrænserne ved hhv. at lægge dette tal til og trække det fra gennemsnittet.

Bakteriæmitallene fra før er: 29, 18, 32, 26, 21, 28, 30, 17, 27, 30, 26, 19, 19, 26, 27, 27, 26, 35, 24, 28, 27, 21, 17, 24. Gennemsnittet er 25.2, og den gennemsnitlige absolutte parvise forskel er 5.78. I-diagrammets kontrolgrænser er derfor $25.2 \pm 2,66 \times 5.78 = 9.8 - 40.6$.

Indikatorer

En indikator er som nævnt et tal, som er udtryk for kvaliteten af et produkt eller en tjenesteydelse, og som kan gøres til genstand for statistisk analyse. Traditionelt opdeler man indikatorer i struktur-, proces- og resultatindikatorer.

Strukturindikatorer knytter sig til fysiske og organisatoriske rammer, fx: findes der en brandindstruks? (ja/nej), hvor mange kvadratmeter er der per sengeplads? (decimaltal), eller hvor mange håndvaske findes på afdelingen? (antal). Strukturindikatorer er i sagens natur ofte relativt statiske og bliver derfor sjældent brugt direkte til løbende kvalitetsudvikling. Men strukturindikatorer kan være nyttige til at forstå rammerne og mulighederne for fremtidig kvalitetsudvikling.

Procesindikatorer knytter sig til rutiner og arbejdsgange, fx: ved hvor stor en andel af operationerne har vi gennemført sikker kirurgi-tjeklisten? (procent), eller hvad er den mediane tid fra udskrivelse til afsendelse af epikrise? (tid). Procesindikatorer er kvalitetsudviklerens vigtigste arbejdsredskab, fordi de siger noget om den adfærd, som i sidste ende har betydning for slutresultatet.

Resultatindikatorer knytter sig til kvaliteten set med kundens øjne, fx: hvor mange hospitalsinfektioner har vi? (antal), hvad er overlevelsen efter operation? (procent), eller i hvilken grad hjælper behandlingen? (decimaltal på VAS-skala eller udsagn på Likert-skala). For kunden er resultatindikatorer naturligvis det mest interessante. Men for kvalitetsudvikleren er resultatet blot det naturlige produkt af strukturer og processer. Resultatet bliver ikke bedre af at blive målt, men af grundlæggende forandringer i strukturer og processer.

Endelig er der **ulempeindikatorer**, som i sig selv kan være resultat- eller procesindikatorer. Man bruger ulempeindikatorer til at opdage eventuelle utilsigtede bivirkninger af de forandringer, man indfører. Det kunne fx være genindlæggelser efter indførelse af accelererede patientforløb. Det hjælper jo ikke, at man reducerer indlæggelsestiden, hvis det medfører, at patienterne kommer igen med komplikationer på et senere tidspunkt. Dette kan ulempeindikatoren være med til at følge.

Afhængig af deres formål kan nogle indikatorer optræde som både den ene og den anden type. I et projekt om forebyggelse af medicinrelaterede genindlæggelser, kan 'andel epikriser sendt inden to hverdage' opfattes som en procesindikator, fordi afsendelse af epikriser skal medvirke til at nå målet. Men den vil kunne opfattes som en resultatindikator, hvis projektets formål er forøgelse af andel epikriser sendt inden for to hverdage.

Måletal og tælleletal

Som det fremgår af eksemplerne ovenfor, findes indikatorer i mange former med det til fælles, at de kan udtrykkes i tal. Indikatorens talformat har betydning for, hvordan den skal opgøres og præsenteres i fx kontroldiagrammer. Grundlæggende skelner man mellem måle- og tælleletal.

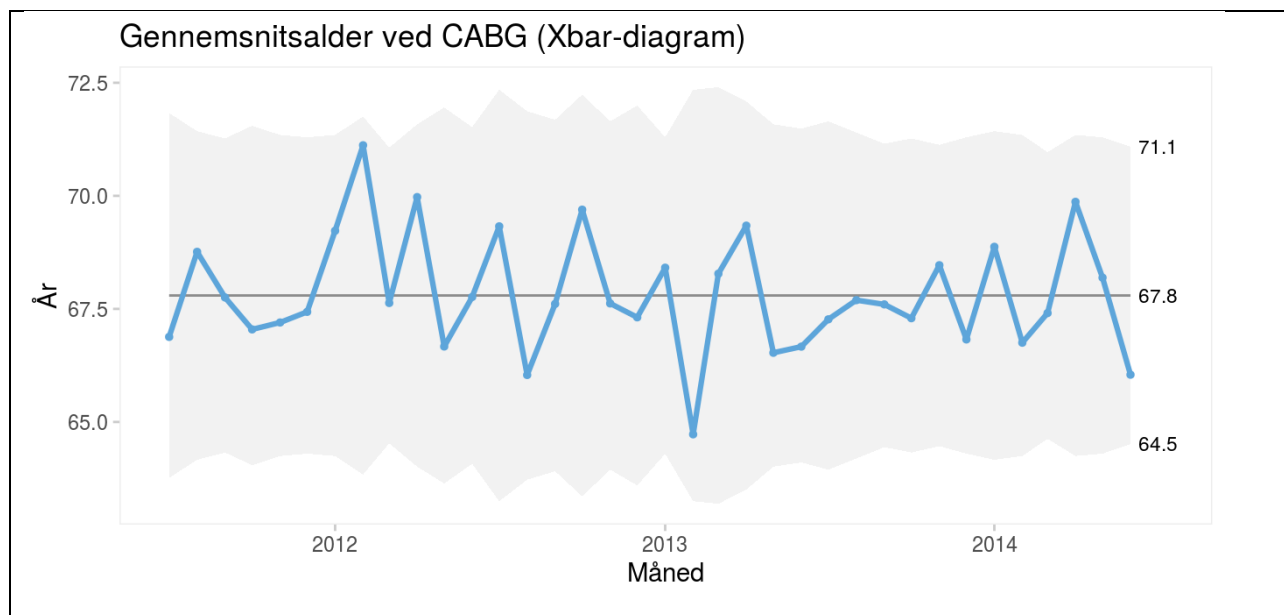
Måletal kendes ved at kunne udtrykkes på en kontinuert skala, som i princippet strækker sig fra $-\infty$ til ∞ , og hvor værdierne kan have uendeligt mange decimaler. I praksis har mange måletal dog naturlige grænser. Fx kan fysiologiske værdier som højde, vægt, puls, blodtryk osv. i sagens natur ikke være negative.

I produktionsindustrien benytter man i vid udstrækning måletal til kvalitetsudvikling og -kontrol. Fysiske produkter som skruer, bolte og bordplader forventes at leve op til præcise standarder for mål og vægt. Til disse formål vil man derfor bruge kontroldiagrammer til måletal: Et I-diagram, hvis det enkelte målepunkt stammer fra en enkelt måling, eller et

Xbar-diagram, hvis et målepunkt er gennemsnittet af flere målinger taget på samme tid eller fra samme stikprøve.

I sundhedsvæsenet er måletal relativt sjældne i forbindelse med kvalitetsudvikling, fordi sundhedskvalitet oftest handler om antal og andele, som er tælleletal. Undtagelserne er tid, fx tid til afsendelse af epikrise, og fysiologiske parametre hos enkeltpersoner, fx alder, vægt, blodtryk peak flow, blodsukker, INR osv.

Diagrammet i figur 6 viser gennemsnitsalderen hos patienter, som får foretaget koronar bypassoperation.

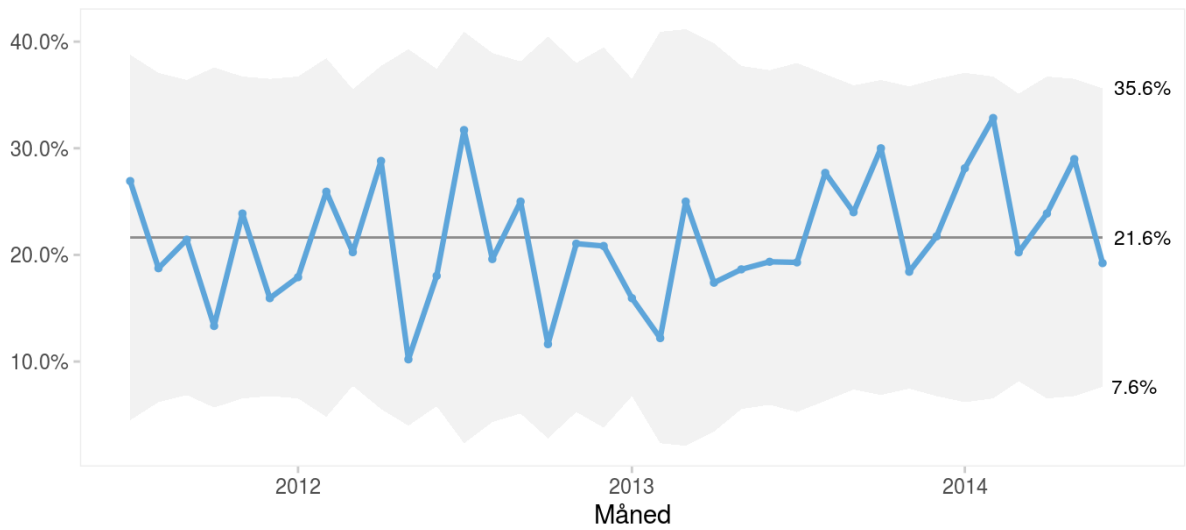


Kontrolgrænserne varierer fra måned til måned fordi, nævneren (antal operationer) varierer. På den måde tager diagrammet højde for naturlige udsving i den almindelige variation, når nævneren varierer. Det samme gælder P- og U-diagrammer.

Tælleletal opstår, når man tæller enten individer med bestemte egenskaber (defekte) eller forekomsten af bestemte fænomener (defekter). Tryksår er et eksempel på et fænomen, der kan tælles som både defekte og defekter: antal patienter med tryksår eller antal tryksår. Tæller man antal patienter med tryksår, tæller en patient kun én gang uanset, hvor mange tryksår vedkommende har. Tæller man antal tryksår, tæller hvert tryksår for sig. I begge tilfælde kan man vælge at udtrykke tallet i relation til noget andet, udfaldsrummet, som så indgår som nævner i brøken. Patienter med tryksår kan fx udtrykkes som andelen af patienter med tryksår, og tryksår kan fx udtrykkes som antallet af nyopståede tryksår per 1000 sengedage. Det første tal er en proportion (andel, procent), det andet tal er en rate. Man kender forskel på proportioner og rater ved, at proportioner altid har samme enhed i tæller og nævner (antal patienter med tryksår/antal patienter i alt), og at tælleren ikke kan være større end nævneren, mens rater har forskellige enheder i tæller og nævner (antal tryksår/antal sengedage).

Det er vigtigt at kende forskel på proportioner og rater, når man skal vælge kontroldiagram. Proportioner analyseres med P-diagrammer (P for proportion eller procent). P-diagrammet i figur 7 viser procentdelen af opererede patienter, som genindlægges inden for 30 dage efter operation.

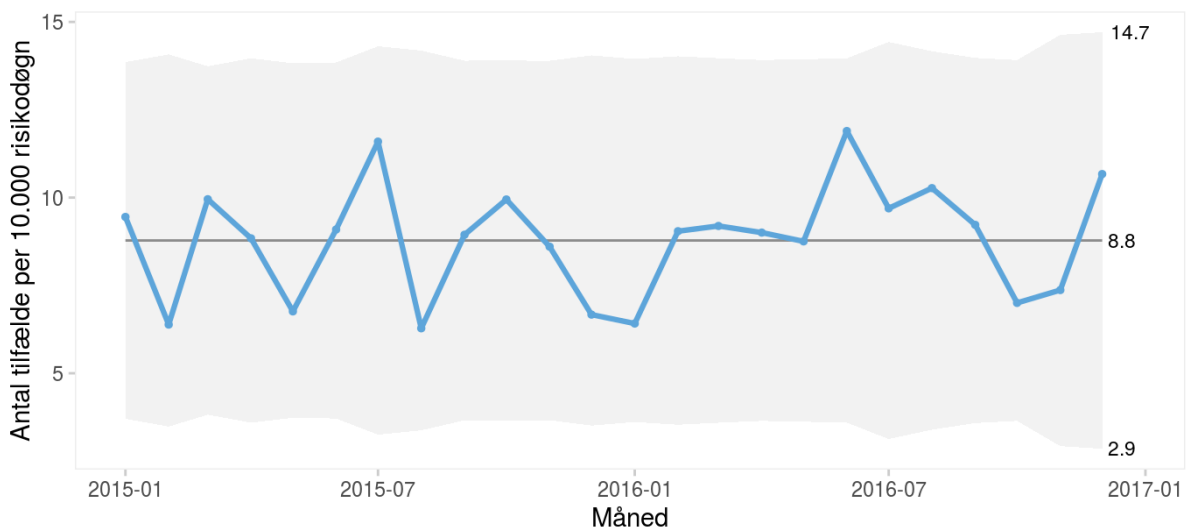
Genindlæggelser efter bypass-operation inden for 30 dage (P-diagram)



Figur 7: P-diagram over procentdelen af genindlæggelser efter bypass

Rater analyseres med U-diagrammer (U for *unequal area of opportunity*).

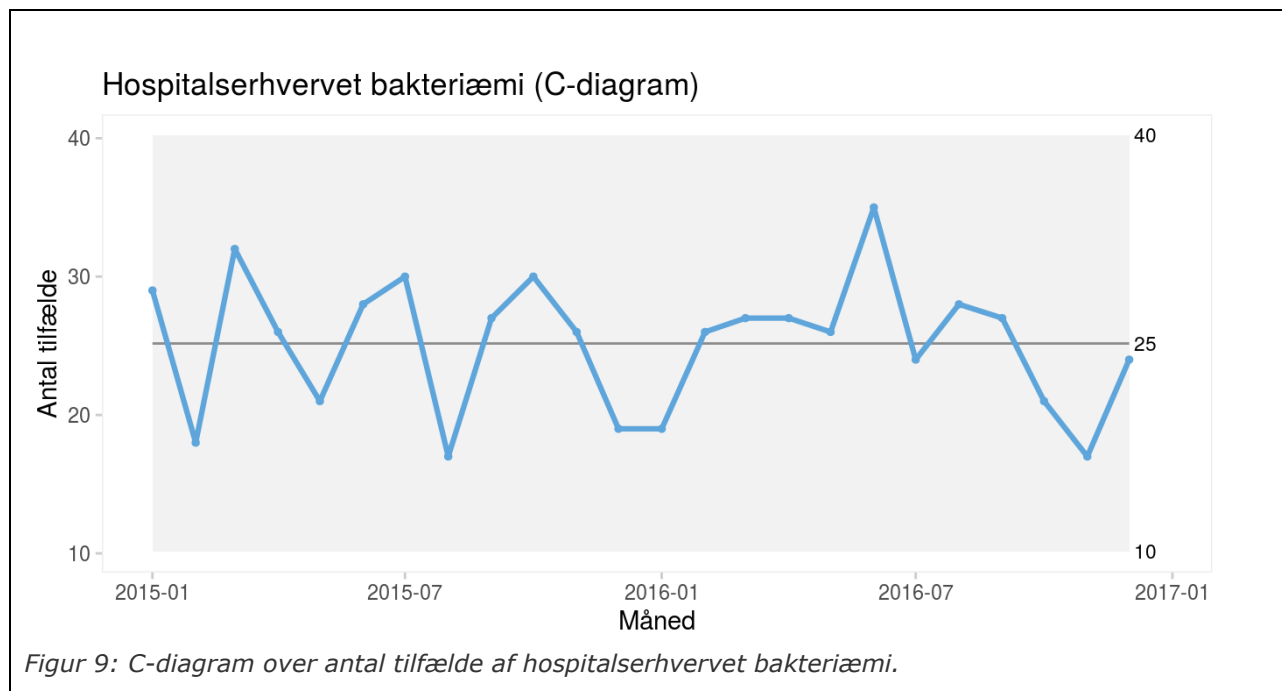
Hospitalserhvervet bakteriæmi (U-diagram)



Figur 8: U-diagram over antal tilfælde af hospitalserhvervet bakteriæmi.

Hvis ratens udfaldsrum (*area of opportunity*) er nogenlunde konstant fra måling til måling, fx antal risikodøgn fra måned til måned som i diagrammet ovenfor, kan man udelade nævneren og benytte et C-diagram (C for *count*). Fordi C-diagrammer viser et antal frem for en rate, kan de være lettere at forholde sig til end U-diagrammer. Således vil de fleste

nok finde det lettere at forholde sig til 25 bakteriemier om måneden end 8,8 bakteriemier per 10.000 risikodøgn.



Bemærk i øvrigt, at C-diagrammet (figur 9) har samme gennemsnit og kontrolgrænser (efter afrunding) som I-diagrammet (figur 2 – om bakteriemier i afsnittet om Shewharts kontroldiagram) med de samme data, vi tidligere beregnede "i hånden".

Stramme indikatorer

Kvalitetsudviklerens vigtigste råstof er gode indikatorer, som kommer af gode data og gode indikatordefinitioner.

Der kan siges og skrives meget om "gode data". Om de er gode, afhænger i virkeligheden af, hvad data skal bruges til. Ønsker vi bare at vide, om vi har elimineret tryksår eller ej, skal vi blot registrere det første tryksår, der opstår. Men hvis vi fx vil vide, om risikoen for tryksår på vores afdeling er signifikant højere eller lavere end på en tilsvarende naboafdeling, skal vi indsamle data for langt flere patienter med og uden tryksår over lang tid, og vi skal være helt sikre på, at vi tæller tryksår på samme måde hele tiden og på samme måde, som de gør på den afdeling, man sammenligne sig med.

Kort sagt: gode data er data, som er *gode nok* til formålet.

Gode indikatordefinitioner kan man derimod ikke snakke sig fra. En indikatordefinition er en klar og utvetydig opskrift på, hvordan man omsætter grunddata til en STRAM indikator:

- **Sensitiv:** indikatoren er følsom for de forandringer, man ønsker at indføre eller undgå.
- **Troværdig:** indikatoren giver et reelt billede af den kvalitet, man ønsker at måle.
- **Reproducerbar:** man når frem til de samme tal og konklusioner, når man gentager analysen på de samme grunddata.
- **Aktuel:** indikatoren viser, hvordan kvaliteten er her og nu – ikke sidste år eller på naboafdelingen.

- **Meningsfuld:** indikatoren giver mening for brugerne.

De færreste indikatorer er det hele. Tag fx den nu hedengange nationale indikator for hospitalsstandardiseret mortalitetsrate, HSMR. HSMR udtrykker forholdet mellem den faktiske dødelighed og den forventede dødelighed på et hospital, hvor der er taget højde for patientsammensætningen (alder, køn, diagnoser, komorbiditet osv.). HSMR er både meningsfuld, aktuel og reproducerbar. Men den er også – viste det sig – meget lidt sensitiv for reelle kvalitetsforbedringer og -forværringer på hospitalerne. Til gengæld er den følsom for meget andet, fx ændringer i lokal registreringspraksis.

Et andet eksempel er overholdelse af udredningsretten. Indikatoren angiver, hvor stor en andel af patienterne, som har en diagnose eller en plan inden for en fastsat tidsramme. Den er sensitiv, aktuel, reproducerbar og troværdig. Men, fordi den stiller alle symptomer og fund lige, er den i mange tilfælde meningsløs for både klinikere og befolkning.

Et eksempel på indikatorer, som er mindre troværdige, fordi de ikke måler præcis det, man ønsker at måle, men alligevel er nyttige, er forekomsten af hospitalserhvervede infektioner. Hospitalsinfektioner er infektioner, som opstår efter hospitalskontakt (typisk mindst 48 timer efter indlæggelse). I opgørelsen af hospitalsinfektioner indgår oplysninger om patient, indlæggelsestidspunkt og tidspunkt for påvisning af infektion, fx fund af bakterier i urin eller blod. Det giver sig selv, at indikatoren ikke er særligt præcis. Nogle patienter indlægges med infektioner, som først påvises efter et par dage og derfor fejlagtigt tælles med (falsk positive); og andre patienter får en hospitalsinfektion allerede i løbet af de første indlæggelsesdøgn og tælles derfor ikke med (falsk negative). Indikatoren kan altså ikke bruges til at fortælle præcis, hvor mange hospitalsinfektioner der reelt forekommer. Men fordi unøjagtighederne er jævnt og tilfældigt fordelt, kan skift i indikatorniveauet over tid alligevel tages som udtryk for forandringer, som i det mindste bør undersøges nærmere.

Arbejdet med at udvikle gode indikatordefinitioner, som giver STRAMme indikatorer, kan erfaringsmæssigt være noget af det mest tidskrævende og vanskelige inden for klinisk kvalitetsudvikling. Men det skal ikke afholde en fra at gå i gang. Og så skal man huske, at det inden for kvalitetsudvikling er tilladt – ja, endda ønskværdigt – at forbedre sine definitioner undervejs, efterhånden som man bliver klogere.

Der er skrevet tykke bøger om indikatorer; men det væsentlige er, at man har en struktureret tilgang til indikatordefinition, så man 'når hele vejen rundt om' indikatoren, inden man går i gang med at samle data

Her er et eksempel på definition af en procesindikator til overvågning af andelen af pneumonipatienter, som har modtaget første dosis antibiotika inden for fire timer efter ankomst.

- **Indikatornavn:** Rettidig diagnostik og behandling af samfundserhvervet pneumoni.
- **Type:** Procesindikator.
- **Formål:** At reducere dødeligheden hos indlagte med samfundserhvervet pneumoni.
- **Tællerdefinition:** Antal pneumonipatienter, som har modtaget første dosis antibiotika inden for fire timer efter ankomst.
- **Nævnerdefinition:** Antal pneumonipatienter.
- **Datakilde:** Morgenkonference.
- **Dataindsamling og -behandling:** Ved morgenkonferencen noteres (for hver nyindlagt patient med samfundserhvervet pneumoni i det forudgående døgn) dato og tidspunkt for indlæggelse og for første dosis antibiotika. Rådata noteres i et

regneark, som automatisk udregner differencen i timer mellem ankomst og behandlingsstart og derefter udregner den ugentlige andel af patienter, hvor antibiotikabehandling er påbegyndt rettidigt.

- **Opgørelsesperiode:** Ugentligt.
- **Analyse og præsentation:** Seriediagram.

Særligt punktet om dataindsamling og -behandling kræver afprøvninger og omhyggelighed, så arbejdet med at indsamle data ikke kommer til at overskygge selve forbedringsindsatsen. I praksis kan man med fordel angibe dataindsamlingsdelen som et lille selvstændigt udviklingsprojekt efter alle kunstens regler med formulering af målsætning og iterativ afprøvning af ideer efter PDSA-metoden.

Det er også vigtigt at være omhyggelig med formulering af tæller- og nævnerdefinitioner, så det er helt tydeligt, hvad der skal inkluderes det ene og det andet sted, så Bente og Børge tæller på samme måde mandag morgen og lørdag nat.

Tip til praktisk dataindsamling og -behandling

I eksemplet med pneumonipatienter indsamler man data manuelt ved morgenkonferencen. Men hvis data allerede er tilgængelige i fx den elektroniske patientjournal, kan opgaven måske automatiseres. Bemærk, at data i dette tilfælde ikke behøver være personhenførbare, idet det er processerne omkring diagnostik og behandling, der er interessante – ikke den enkelte patient.

Når man samler data, skal man så vidt muligt undgå at bearbejde data allerede i indsamlingsøjeblikket. Man kunne i eksemplet have valgt blot at notere for hver patient, om behandlingen var rettidigt indledt (ja/nej). På den måde kunne dataindsamlingen måske gøres lidt enklere. Men samtidig ville man afskære sig selv for på et senere tidspunkt at opgøre data på andre måder.

Ved at dikotomisere grunddata, som det hedder, når man oversætter en kontinuert variabel (i dette tilfælde tid) til en logisk variabel (ja/nej), fjerner man information. Hvis man derimod indsamler grunddata så ubearbejdet som muligt, bliver det muligt sidenhen at opgøre data på flere forskellige måder. Det kan være, at ny evidens på et tidspunkt gør det interessant at vide, hvilken andel af patienterne, som kom i behandling inden for to, otte eller syvtrekvarter timer. Det kan også være, man ønsker at se på tid til behandlingsstart for individuelle patienter eller den mediane eller gennemsnitlige tid for alle patienter per uge. Alle disse muligheder går tabt i det øjeblik, man allerede i dataindsamlingsøjeblikket reducerer data.

En beslægtet kunstfejl er at aggregere data allerede ved kilden. Dette sker ofte, når man samler data ved auditmetoden, hvor man gennemgår et antal patientforløb for en given tidsperiode. Det er fristende blot at notere tæller og nævner for hele perioden. Men ved at notere grunddata for det enkelte forløb, mangedobler man sine muligheder for at opgøre indikatoren på forskellige måder. I eksemplet ovenfor aggregerer man data efter, de er indsamlet. På den måde er det let, siden hen at ændre fx opgørelsesperiode fra uge til måned.

Særligt gælder det for tidsintervaller, at de kan være vanskelige eller umulige at ændre. Man kan ikke lave månedsopgørelser med ugedata og omvendt. Men hvis grunddata indeholder de eksakte datoer for det, man undersøger, kan man opgøre data i alle tænkelige tidsintervaller: dag, uge, to-, tre-, fire-ugersperiode, måned, kvartal osv.

Dette betyder dog ikke, at man *aldrig* skal reducere eller aggregere data allerede i indsamlingsøjeblikket. Der kan være særdeles gode grunde til at gøre det. Men man skal

være bevidst om sin indsamlingsstrategi og nøje overveje fordele og ulemper ved den ene og den anden metode. En god grund til at aggregere data – ud over at lette dataindsamlingen – kan være at maskere personhenførbare data. På den måde slipper man for bøvl med tilladelser og krav til databasesystemer med logning osv. Særligt i tidsbegrænsede her og nu-projekter kan det betyde forskellen på succes og fiasko, om man skal vente på tilladelser til at gå i gang.

Sammenfattende kan man sige, at man skal tilstræbe så høj opløsning som muligt i sine grunddata under hensyntagen til dataindsamlingsressurser og behovet for maskering af personhenførbare oplysninger.

Case: Tid til grad 2-sectio

På fødeafdelingen på Hospitalsenhed Vest (Herning og Holstebro) arbejder man bl.a. med forløbstider ved akutte kejsersnit (case beskrevet i kapitel 11) Efter hvert kejsersnit noterer jordemoderen fødselstidspunkt, tidspunkt for ordination af kejsersnit samt i hvilken grad kejsersnittet var akut (grad 1-3). For hver akuthedsgrad er der en tidsgrænse, som man tilstræber overholdt for mindst 90% af kejsersnittene. For grad 2-kejsersnit er grænsen 30 minutter.

Data omfatter i alt 354 grad 2-kejsersnit fra perioden januar 2016 til august 2019. Tabellen viser data for de første 6 kejsersnit.

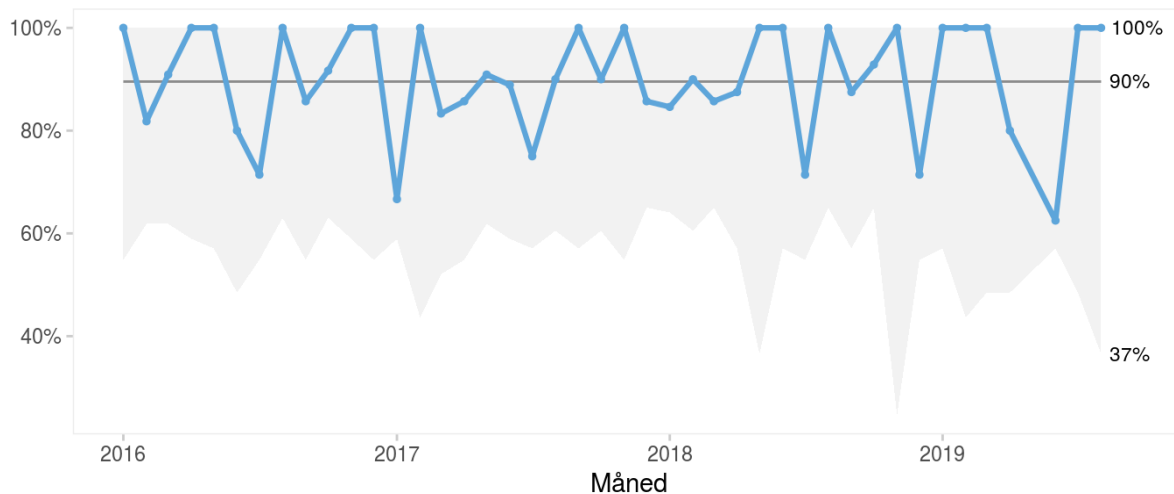
datetime	month	sectio_delay (minutter)
2016-01-06 03:55	2016-01-01	22
2016-01-06 20:52	2016-01-01	22
2016-01-07 02:50	2016-01-01	29
2016-01-07 22:32	2016-01-01	28
2016-01-09 14:56	2016-01-01	22
2016-01-09 21:21	2016-01-01	20

Variablen datetime angiver fødselstidspunktet; month er fødselsmåneden, som er beregnet ud fra datetime. Vi skal bruge fødselsmåneden til at aggregere data til P- og Xbar-diagrammer; sectio_delay er tiden (i minutter) fra ordination af kejsersnit til barnet er født. Af hensyn til beskyttelse af personoplysninger er der tilføjet støj til datetime-variablen, så fødselstidspunkterne i tabellen afviger fra de faktiske fødselstidspunkter.

For di grunddata registreres for hvert enkelt kejsersnit, er det muligt at tilpasse sine opgørelser til forskellige målgrupper.

Hospitals- og regionsledelsen ønsker måske blot at vide, om standarden bliver overholdt, og om overholdelsen er stabil. P-diagrammet viser, at 90% af kejsersnittene sker inden for tidsrammen, og at processen er stabil.

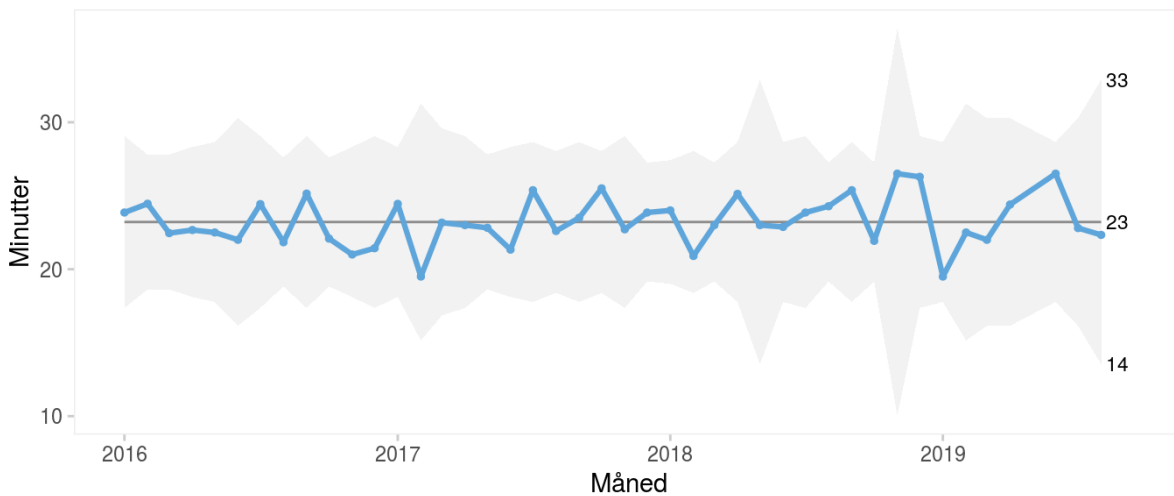
Grad 2-kejsersnit gennemfør inden for 30 minutter efter ordination



P-diagram

Afdelingens læger og jordemødre ønsker måske at kende den gennemsnitlige tid fra ordination af kejsersnit til fødsel. Det viser Xbar-diagrammet, hvor hvert målepunkt angiver den gennemsnitlige tid til kejsersnit i hver måned. Alle målepunkter ligger inden for kontrolgrænserne og omkring en gennemsnitstid på 23 minutter.

Gennemsnitlig tid til gennemførelse af grad 2-kejsersnit



Xbar-diagram

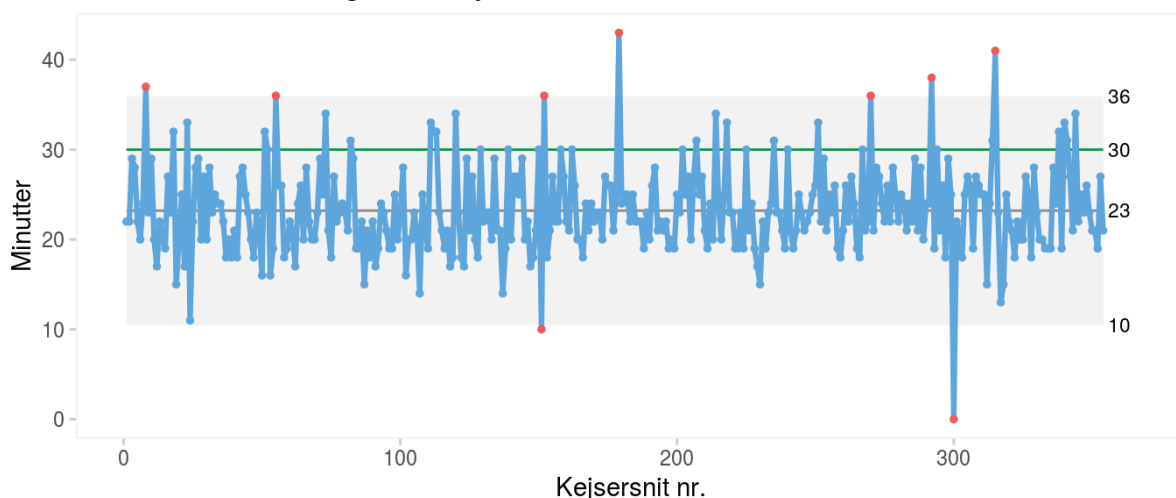
Data til P- og Xbar-diagrammerne fås ved at aggregere grunddata per måned. Tabellen ovenfor viser aggregerede data for de første seks måneder. Variablen n_{ontime} angiver antallet af grad 2-kejsersnit, som er udført inden for tidsgrænsen; n_{sectio} angiver det totale antal grad 2-kejsersnit; p_{ontime} angiver proportionen af kejsersnit udført til tiden og udgør datapunkterne i P-diagrammet; avg_delay angiver den gennemsnitlige tid til kejsersnit og udgør datapunkterne i Xbar-diagrammerne; og sd_delay angiver standardafvigelsen i tid til kejsersnit. Kontrolgrænserne i de to diagrammer udregnes ved at indsætte de relevante parametre i formlerne for hhv. P- og Xbar-diagrammet.

month	n_ontime	n_sectio	p_ontime	avg_delay	sd_delay
2016-01-01	7	7	1.00	23.86	3.39
2016-02-01	9	11	0.82	24.45	6.14
2016-03-01	10	11	0.91	22.45	6.64
2016-04-01	9	9	1.00	22.67	3.04
2016-05-01	8	8	1.00	22.50	3.89
2016-06-01	4	5	0.80	22.00	6.20

Forbedringsteamet, hvis opgave det er at skabe robuste processer og arbejdsgange, som sikrer den enkelte fødende højst mulig kvalitet, har brug for flere detaljer. Dem finder de i et I-diagram over individuelle forløbstider, hvor hvert målepunkt er forløbstiden for ét enkelt kejsersnit. I-diagrammet har (naturligvis) det samme gennemsnit på 23 minutter som Xbar-diagrammer, men viser tillige, at processen er ustabil på individniveau, idet enkelte kejsersnit tager usædvanligt lang eller kort tid.

I-diagrammet viser også, at den nuværende proces ikke tillader 100%-gennemførelse inden for tidsgrænsen. Det fremgår ved, at øverste kontrolgrænse på 36 minutter er over tidsgrænsen på 30 minutter, som er angivet med den vandrette grønne streg.

Individuelle tider til grad 2-kejsersnit



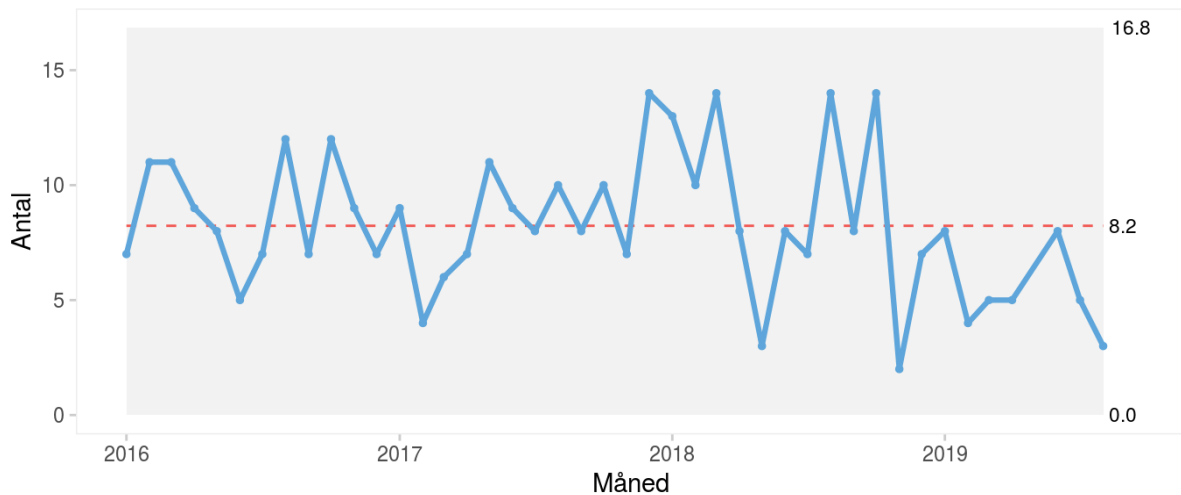
I-diagram

Hvert diagram har sine styrker og svagheder, og de tre diagrammer viser tre sider af de samme data og besvarer i virkeligheden forskellige spørgsmål. Til administratorer og til offentligheden kan vi sige, at vi overholder standarden og holder et stabilt niveau, dvs. det går godt. Til os selv kan vi sige, at der er plads til forbedring i form af stabilisering af individuelle forløbstider. Hvis vi kan identificere årsagerne til de kejsersnit, som befinder sig uden for kontrolgrænserne og dermed skiller sig ud på den gode eller den dårlige måde, kan vi udover at stabilisere processen måske endda reducere den gennemsnitlige forløbstid.

Endelig kunne vagtplanlæggeren af hensyn til planlægning af den fremtidige bemanning i vagterne måske ønske at vide, om antallet af kejsersnit er stabilt og forudsigeligt fra måned

til måned. Dette spørgsmål kan besvares med et C-diagram over det månedlige antal kejsersnit.

Grad 2-kejsersnit



C-diagram

C-diagrammet viser særlig variation i form af et skift med en usædvanligt lang serie på ni datapunkter mod forventet højst otte til sidst på kurven. Noget kunne altså tyde på, at der er begyndt at komme færre kejsersnit. Dette bør undersøges nærmere. Der kan være mange forklaringer. En forklaring kunne være, at faldet skyldes bevidste tiltag for at reducere hyppigheden af kejsersnit.

Dette ville formentlig være en positiv udvikling. Det kunne også være, at man af den ene eller anden grund var begyndt at foretage flere elektive kejsersnit og dermed færre akutte, eller at fødselstallet simpelthen var faldet. Man må overlade det til fagfolk med godt lokalkendskab at afgøre, om skiftet er ønsket, uønsket eller ligegyldigt. Uanset hvad, vil det være fornuftigt at lede efter en forklaring.

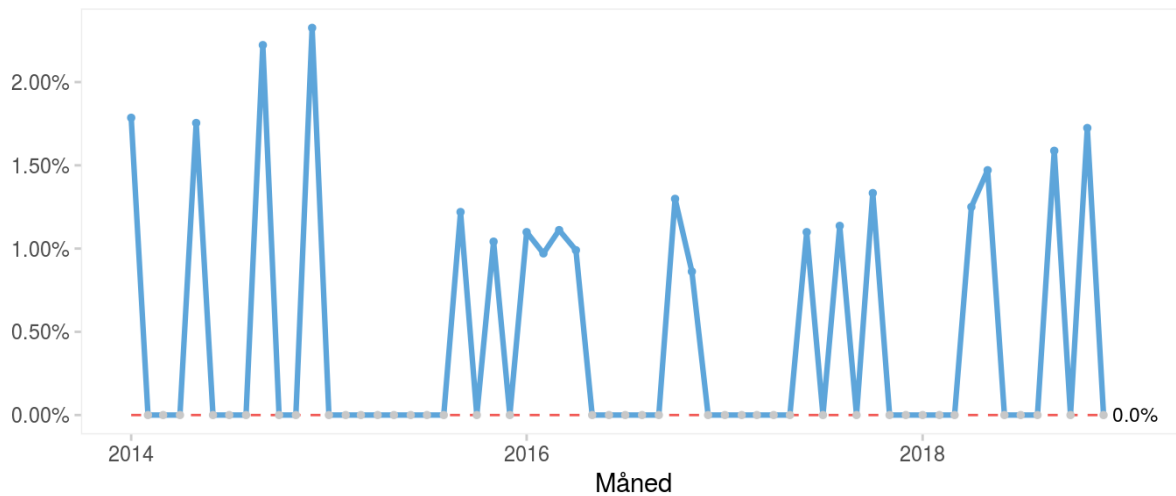
Case: Sjældne kirurgiske infektioner

Vi kalder det sjældne hændelser, når det, vi tæller, sker så sjældent, at vi ikke kan samle nok data til, at SPC-diagrammet giver mening. Det viser sig ved, at selv måneds- eller kvartalsopgørelser (for) ofte indeholder nul defekte(r).

"Sjælden" er et relativt begreb. På hospitalsniveau er hospitalsinfektioner sjældent sjældne – en om dagen på Rigshospitalet. Men den lille afdeling eller den enkelte læge oplever måske kun en enkelt eller slet ingen hospitalsinfektioner på et helt år.

I seriediagrammet nedenfor er medianen nul, fordi mere end halvdelen af målepunkterne er nul. Derfor giver det ikke mening at lede efter lange serier og få kryds; der er kun én serie og ingen kryds.

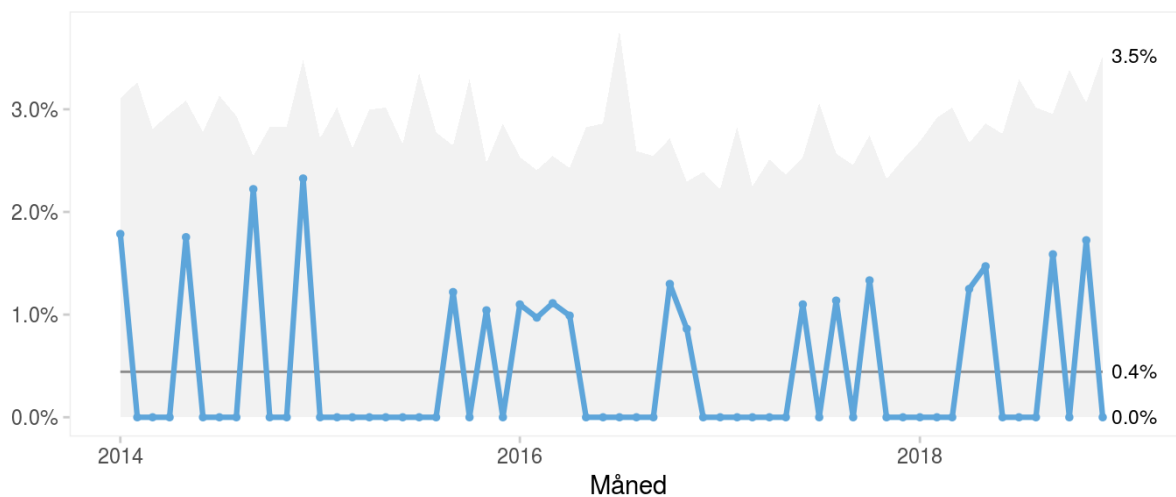
Dybe infektioner inden for 90 dage efter hofteprotese



Kilde: haiba.dk, Herlev-Gentofte Hospital

P-diagrammet er mere brugbart, men tallene er stadig så små og kontrolgrænserne derfor så brede, at det ville tage lang tid for os at opdage hvis infektionshyppigheden skulle ændre sig markant. Der skal ske en mangedobling af infektionsrisikoen, før kurven krydser den øvre kontrolgrænse. Og, fordi nedre kontrolgrænse er censureret ved nul, skal der gå mange måneder uden infektioner, før vi opdager en reduktion i infektionsrisikoen.

Dybe infektioner inden for 90 dage efter hofteprotese (P-diagram)



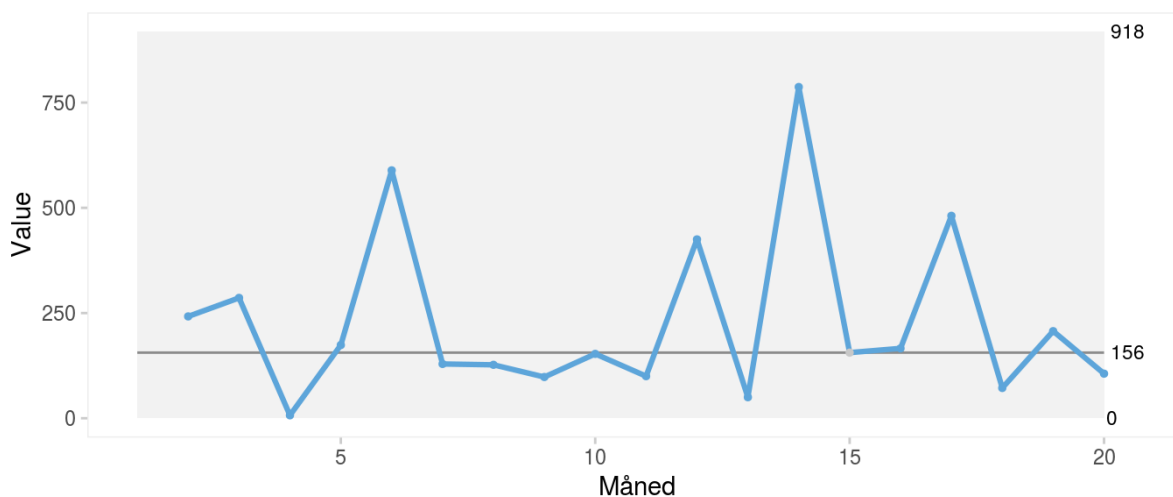
Kilde: haiba.dk, Herlev-Gentofte Hospital

Når hændelser bliver så sjældne, at de ikke kan opgøres meningsfuldt i passende tidsintervaller på x-aksen, kan man i mange tilfælde slippe af sted med at omdefinere indikatoren – vende den om så at sige – så man i stedet for at tælle antal hændelser per tidsenhed tæller tiden mellem hændelser. På afdelingen kan man fx tælle antal dage mellem nyopståede infektioner eller andre uønskede hændelser. Kirurgen vil måske tælle antal operationer mellem postoperative komplikationer. Og medicineren kan tælle, hvor mange patienter, der overlever et AMI mellem hvert AMI-relateret dødsfald.

Til disse typer indikatorer findes såkaldte T- og G-diagrammer. T for tid mellem hændelser (fx dage mellem infektioner) og G for ikke-defekte enheder mellem defekte enheder (fx operationer mellem komplikationer) (Benneyan 2002).

G-diagrammet nedenfor viser, at der mediant går 156 operationer mellem, der opstår en infektion, og at en sikker forbedring (= færre infektioner) vil vise sig den dag, der går op mod 1000 operationer mellem infektioner.

Operationer mellem dybe infektioner efter hofteprotektkirurgi (G-diagram)



Kilde: haiba.dk, Herlev-Gentofte Hospital

Bemærk, at fordi G-diagrammet baserer sig på antagelser om, at data følger den geometriske sandsynlighedsfordeling, som er stærkt asymmetrisk, bruger man som regel medianen som midtlinje. På den måde kan man stadig applicere test for lange serier og få kryds, som i sagens natur forudsætter at data er nogenlunde symmetrisk fordelt om midtlinjen.

G-diagrammet er dog ingen mirakelkur mod sjældne hændelser. Bemærk, at der skal gå op til 1000 operationer mellem infektioner, før vi krydser øvre kontrolgrænse. Med i gennemsnit 75 operationer om måneden, betyder det, at der skal gå over et år uden infektioner, før vi kan være sikre på, at risikoen reelt er reduceret. I P-diagrammet ovenfor skal vi kun vente ni måneder, før vi har en usædvanlig lang serie under midtlinjen.

Man skal også være opmærksom på, at G-diagrammet sjældent har nogen nedre kontrolgrænse og derfor er uegnet til at opdage forværringer, som viser sig ved, at kurven går nedad.

I virkeligheden supplerer P- og G-diagrammerne hinanden, og man kan med fordel vise dem sammen. Det er min erfaring, at det er hip som hap, hvilket af de to, som først signalerer, når processen skifter.

Håndværk, Videnskab og Tankegang

SPC er på samme tid et håndværk og en videnskab – men først og fremmest er SPC en tankegang (Wheeler 2000). Håndværket og videnskaben kan man træne og læse sig til. Det svære er at tilegne sig tankegangen.

SPC-tankegangen er tydeligst udtrykt i Tabel 1 med de fire tilstande og tilhørende handlinger. Den kan man jo hurtigt læse op på. Men det kræver tid og øvelse at

internalisere den. Til gengæld er udbyttet besværet værd. For den kvalitetsbevidste organisation er SPC det sunde alternativ til skiftende bevidstløse og skadelige indgreb i stabile processer og blind tiltro til ustabile processer på afveje.

SPC har siden sin fødsel for snart 100 år siden revolutioneret produktionsindustrien verden over. I sundhedsvæsenet og andre servicebrancher er det gået noget langsommere. Selv om mange klinikere og kvalitetsmedarbejdere har taget SPC til sig, kan det tælles på et par fingre de steder, hvor SPC er integreret fra top til bund i hele ledelses- og produktionssystemet.

Vi håber, at dette kapitel kan være med til at udruste fremtidens klinikere og ledere, så SPC får den plads i sundhedsvæsenet, som patienterne (det er også os) fortjener.

Hyppigt stillede spørgsmål

Hvor mange data er nødvendige til et SPC-diagram?

Kort sagt: der går 20-30 målepunkter til et godt SPC-diagram. Er der færre risikerer man i højere grad at overse særlig variation. Er der flere risikerer man flere falske signaler. Dette betyder ikke, at flere eller færre målepunkter ikke kan bruges. Man skal naturligvis udnytte de data, man har, blot man er opmærksom på betydningen af antallet af datapunkter for fortolkningen.

I kejsersniteksemplet ovenfor er der 354 datapunkter i I-diagrammet. I fortolkningen bør man derfor være opmærksom på den forøgede risiko for falske signaler, som findes, når der er så mange datapunkter. I eksemplet kan det måske betale sig at "nøjes" med at undersøge de tre datapunkter, der ligger længst fra kontrolgrænserne. Der er formentlig i dem, man finder den vigtigste læring om, hvad der en gang i mellem går særligt galt og godt.

Hvor hyppigt skal jeg måle?

Det giver sig selv, at hvis man skal bruge 20-30 datapunkter til et godt SPC-diagram, så er der ikke meget kvalitetsudvikling at hente i års- og kvartalsdata. Selv med månedlige data, vil der gå op til to år, før vi med sikkerhed kan afgøre, om processen er stabil eller ej. Man skal altså måle hyppigt.

En anden god grund til at måle hyppigt er det, der i fagsproget kaldes *rational subgrouping*, som betyder noget i retning af fornuftig inddeling af x-aksen. For at maksimere chancerne for at opdage særlig variation, skal man måle hyppigt nok til, at eventuelle skift vil forekomme mellem målepunkterne. Eller med andre ord: målefrekvensen skal være højere end forandringsfrekvensen. Hvis man fx forventer, at forandringer kan ske fra uge til uge, skal man måle dagligt eller i det mindste ugentligt.

Omvendt risikerer man ved hyppige målinger, at tallene i datapunkterne bliver for små. Dette gælder for tællemaal og medfører, at den tilfældige variation (støjen) fra målepunkt til målepunkt bliver større og dermed lettere kommer til at maskere eventuelle signaler. Som en grov tommelfingerregel kan man tilstræbe, at der i gennemsnit er mindst 4-5 defekte eller defekter i tælleren i hvert målepunkt.

Så på den ene side skal man måle hyppigt nok til, at eventuelle forandringer sker mellem målepunkterne og på den anden side så sjældent, at datapunkterne indeholder nok data.

Er det nødvendigt at have baseline-data før et forbedringsprojekt går i gang?

Det bedste udgangspunkt for et kvalitetsudviklingsprojekt er en god, solid baseline, som med 20-30 målepunkter viser stabil proces. Et godt eksempel er *C. difficile*-casen fra før. I sådanne tilfælde, kan man fastlåse midtlinjen og eventuelle kontrolgrænser og forlænge

dem ud i fremtiden. Når midtlinje og grænser er låst, vil man hurtigere opdage eventuelle skift i data.

Men hvad gør man, hvis man ikke har data til en god, solid baseline? Skal man vente med at forbedre kvaliteten, til man har etableret en baseline, eller skal man bare gå i gang? Svaret kommer an på situationen.

Hvis man har et erkendt og alvorligt problem i sin afdeling, kan man selvfølgelig ikke vente halve eller hele år på en baseline. Så må man i gang med at forbedre kvaliteten med det samme, og indikatormålingerne må følge med, så godt de kan. Det er min erfaring, at de færreste udviklingsprojekter virker med det samme. Det betyder, at man alligevel næsten altid opnår en form for baseline i løbet af det første stykke tid, mens man afprøver sine forandringsideer.

Omvendt, hvis man blot har en fornemmelse af, at noget kunne gøres bedre, uden at det er livet om at gøre, at det sker her og nu, kan man måske godt forsvare at tage den nødvendige tid til at belyse problemet og udviklingsmulighederne grundigt med en SPC-analyse af en omhyggeligt udført baselineregistrering.

Hvornår er det fornuftigt at opdele SPC-diagrammet i flere perioder?

I C. *difficile*-casen (i afsnittet 'Serie- og/eller kontroldiagrammer – hvad skal du vælge?') opdelt vi diagrammet i to perioder før og efter interventionen. På den måde er det let på øjemål at kvantificere forskellen på perioderne. Men før man opdeler sit diagram, skal man sikre sig, at forudsætninger for en opdeling er opfyldt. Når man opdeler et diagram, vil man næsten med garanti få forskellige midtlinjer i de forskellige perioder – også selv om der ikke er sket væsentlige ændringer i de underliggende processer. Det kan snyde læseren til at tro, at der er sket forbedringer (eller forværringer). Denne praksis er udbredt – også i den "finere" litteratur og på videnskabelige kvalitetskonferencer – men den er ikke desto mindre forkert.

Det *kan* være nyttigt at opdele et SPC-diagram når

- der er et tydeligt skift i data,
- årsagen til skiftet er kendt,
- den nye proces er bedre end den gamle, og
- vi forventer, at den nye proces fortsætter.

Hvis blot én af betingelserne ikke er opfyldt, skal man hellere bruge sin tid på at afprøve ideer til arbejdsgange, den kan stabilisere og/eller forbedre processen.

Hvorfor alt det besvær med serie- og kontroldiagrammer?

Kan vi ikke bare sammenligne kvaliteten før og efter en indsats med fx t- eller chi i anden-test?

Sagtens. Blot man sikrer sig, at både før- og efter-perioden hver for sig er statistisk stabile, dvs. kun rummer almindelig variation. Er de ikke det, er forudsætningerne for komparative analyser af gennemsnit og proportioner ikke opfyldt. Det giver ingen mening at sammenligne gennemsnit fra to processer, som er i bevægelse. Det ville svare til at sammenligne gennemsnitshastigheden hos en bil, der accelererer med gennemsnitshastigheden hos en bil, der bremser.

Man kan undersøge før- og efter-perioderne for særlig variation ved at placere dem i et SPC-diagram. Og hvis man gør det, er det slet ikke nødvendigt med t- eller chi i anden-test.

Hvor finder jeg SPC-software?

Man kan komme rigtig langt med hardware i form af papir og blyant. Det kræver få statistiske kundskaber at tegne og fortolke et seriediagram. Og de fleste kan hurtigt lære at udregne kontrolgrænser til et I-diagram.

Men meget bliver naturligvis lettere, hvis man har en computer til at hjælpe sig. Alle de store statistikprogrammer som SAS, JMP, STATA, SPSS og R indeholder moduler, som kan lave kontroldiagrammer. Til Excel kan man købe forskellige *plug ins*, som kan lave SPC-diagrammer. Man kan også relativt let lave sine egne regnearksskabeloner til formålet. Endelig er der EpiData, som er et dansk udviklet statistikprogram, der også kan lave SPC-diagrammer.

Aktuelt (pr. medio 2021) findes skift og kryds-reglerne kun i qicharts2-pakken til R (Anhøj 2018, 2019). De andre programmer benytter forskellige udgaver af Western Electric-reglerne, og de færreste kan lave seriediagrammer. Både qicharts2 og R er *open source* og gratis. De andre programmer er dyre, medmindre man er studerende. Undtagelsen er EpiData, som er gratis til personligt brug, men som man forventes at betale for eller på anden måde bidrage til, hvis man bruger det til arbejds- eller kommercielle formål.

qichart2 forudsætter, at man behersker R til husbehov. R er (ligesom SAS og STATA) et programmeringssprog, som det tager tid at lære. Men har man først lært det, er der ingen grænser for, hvad R i øvrigt kan bruges til – ikke mindst i hænderne på læger, som skal forske og kvalitetsudvikle resten af livet. Kort sagt: Læge, lær R!

Hvor finder jeg gode data til mit kvalitetsudviklingsprojekt?

Det er sjældent, at gode data til kvalitetsudvikling findes på forhånd – i hvert fald i tilstrækkeligt omfang. Man kan finde mange gode data i de elektroniske journalsystemer, i Landspatientregisteret og i de kliniske kvalitetsdatabaser (se også Datakilder, der kan bidrage til analyser af kvaliteten – link til underside til kapitel 2). Men det er sjældent, sådanne data er tilstrækkelige til at drive konkrete kvalitetsudviklingsprojekter.

I eksisterende datakilder finder man især data, som er brugbare til at konstruere resultatindikatorer. Fx stammer rådata om hospitalsinfektioner i ovenstående eksempler fra mikrobiologiske databaser og Landspatientregisteret. Men eksisterende datakilder er næsten altid utilstrækkelige til at konstruere procesindikatorer. I kejsersnitseksemplet samler jordemødrene på Hospitalsenhed Vest selv data ind efter hver fødsel. I journalsystemet findes nemlig ikke pålidelige oplysninger om tidspunkt for ordination af kejsersnit. Faktisk samler jordemødrene mange flere data om den enkelte fødsel til brug for deres lokale kvalitetsarbejde – afhængig af fødselstype noterer de op til over 100 variable i en til formålet oprettet database. De har udviklet en arbejdsgang, hvor denne efterregistrering kan overstås på mindre end fem minutter, hvilket i sig selv er imponerende.

Hvorfor ikke 2 sigma?

Fra al anden statistik har vi vænnet os til, at det normale skal findes inden for +/- 2 sigma fra gennemsnittet. Så hvorfor bruger man 3 sigma i kontroldiagrammer?

Det er let at argumentere statistisk for, at 3 sigma er en bedre afgrænsning af almindelig variation i kontroldiagrammer end 2 sigma. Men læs svaret til ende, før du køber forklaringen.

Kontroldiagrammer er i realiteten gentagne (20-30) statistiske test. Med tilfældige, normalfordelte data er chancen for, at det enkelte datapunkt falder uden for 2 sigma ca. 5%. Men i et kontroldiagram gentager vi testen mange gange. Med 20 datapunkter fra en

stabil proces med normalfordelte data er sandsynligheden for, at *alle* datapunkterne ligger mellem ± 2 sigma = $1 - 0.95^{20} = 0.64$. Det betyder, at hvis vi lader 2 sigma-grænser være afgørende for, om processen er stabil eller ej, skal vi forberede os på, at en tredjedel af vores kontroldiagrammer indeholder et eller flere falske signaler. Med 3 sigma er sandsynligheden for, at en tilfældig måling falder uden for 0.03%, og nu ser regnestykket mere fornuftigt ud: $1 - 0.997^{20} = 0.06$.

Shewhart havde naturligvis selv gjort sig disse overvejelser. Men det er interessant, at han protesterede flere gange i både skrift og tale mod denne statistiske forklaring: *In other words, the fact that the criterion we happen to use has a fine ancestry of highbrow statistical theorems does not justify its use. Such justification must come from empirical evidence that it works.* (Shewhart 1931, p. 18)

Shewhart valgte altså 3 sigma til sine kontrolgrænser af den simple grund, at de fungerer i praksis. Der er i de 100 år, der snart er gået, gjort mange forsøg på at finde bedre kontrolgrænser; men 3 sigma står stadig ubesejret.

Referencer

Anhøj, Jacob. 2015. "[Diagnostic Value of Run Chart Analysis: Using Likelihood Ratios to Compare Run Chart Rules on Simulated Data Series](#)". *PLoS ONE*. .

———. 2018. "[qicharts2: Quality Improvement Charts for R](#)". *JOSS*. .

———. 2019. [qicharts2: Quality Improvement Charts for R](#).

Anhøj, Jacob, og Anne-Marie Blok Hellesøe. 2016. "The problem with red, amber, green: the need to avoid distraction by random variation in organisational performance measures". *BMJ Qual Saf*. .

Anhøj, Jacob, og Anne Vingaard Olesen. 2014. "[Run Charts Revisited: A Simulation Study of Run Chart Rules for Detection of Non-Random Variation in Health Care Processes](#)". *PLoS ONE*.

Anhøj, Jacob, og Tore Wentzel-Larsen. 2018. "[Sense and sensibility: on the diagnostic value of control chart rules for detection of shifts in time series data](#)". *BMC Medical Research Methodology*.

Benneyan, James. 2002. "[Number-Between g-Type Statistical Quality Control Charts for Monitoring Adverse Events](#)". *Health care management science* 4 (januar): 305–18.

Carstensen, Jeppe West, Mahtab Chehri, Kristian Schønning, Steen Christian Rasmussen, Jacob Anhøj, Nina Skavlan Godtfredsen, Christian Østergaard Andersen, og Andreas Munk Petersen. 2018. "Use of prophylactic *Saccharomyces boulardii* to prevent *Clostridium difficile* infection in hospitalized patients: a controlled prospective intervention study". [Eur J Clin Microbiol Infect Dis](#) 37: 1431–9.

Montgomery, Douglas C. 2009. *Statistical Quality Control: A Modern Introduction, sixth ed.* John Wiley & Sons.

[Quality America Inc. 2016. "Tampering". 2016.](#)

Shewhart, Walter A. 1931. *Economic control of quality of manufactured product.* New York: D. Van Nostrand Company.

Western Electric Company. 1956. *Statistical Quality Control Handbook.* New York: Western Electric Company inc.

Wheeler, Donald J. 2000. [*Understanding Variation – The Key to Managing Chaos, third ed.*](#) SPC Press.