



Folkhälsomyndigheten

# Datorstödd utbrottsdetektion

Datorprogrammet CASE – erfarenheter och användarnytta





# Datorstödd utbrottsdetektion

Datorprogrammet CASE – erfarenheter och användarnytta

## **Bindningar och jäv**

För Folkhälsomyndighetens egna experter och sakkunniga som medverkat i rapporter bedöms eventuella intressekonflikter och jäv inom ramen för anställningsförhållandet.

När det gäller externa experter och sakkunniga som deltar i Folkhälsomyndighetens arbete med rapporter kräver myndigheten att de lämnar skriftliga jävsdeklarationer för potentiella intressekonflikter eller jäv. Sådana omständigheter kan föreligga om en expert t.ex. fått eller får ekonomisk ersättning från en aktör med intressen i utgången av den fråga som myndigheten behandlar eller om det finns ett tidigare eller pågående ställningstagande eller engagemang i den aktuella frågan på ett sådant sätt att det uppkommer misstanke om att opartiskheten inte kan upprätthållas.

Folkhälsomyndigheten tar därefter ställning till om det finns några omständigheter som skulle försvåra en objektiv värdering av det framtagna materialet och därmed inverka på myndighetens möjligheter att agera sakligt och opartiskt. Bedömningen kan mynna ut i att experten kan anlitas för uppdraget alternativt att myndigheten föreslår vissa åtgärder beträffande expertens engagemang eller att experten inte bedöms kunna delta i det aktuella arbetet.

De externa experter som medverkat i framtagandet av denna rapport har inför arbetet i enlighet med Folkhälsomyndighetens krav lämnat en deklARATION av eventuella intressekonflikter och jäv. Folkhälsomyndigheten har därefter bedömt att det inte föreligger några omständigheter som skulle kunna äventyra myndighetens trovärdighet. Jävsdeklarationerna och eventuella kompletterande dokument utgör allmänna handlingar som normalt är offentliga. Handlingarna finns tillgängliga på Folkhälsomyndigheten.

---

Denna titel kan laddas ner från: [www.folkhalsomyndigheten.se/publicerat-material/](http://www.folkhalsomyndigheten.se/publicerat-material/).

Citera gärna Folkhälsomyndighetens texter, men glöm inte att uppge källan. Bilder, fotografier och illustrationer är skyddade av upphovsrätten. Det innebär att du måste ha upphovsmannens tillstånd att använda dem.

© Folkhälsomyndigheten, 2017.

Ersätter den tidigare titeln Datorstödd utbrottsdetektion på Smittskyddsinstitutet.

Artkelnummer: 02922-2017.

# Förord

CASE (Computer Assisted Search for Epidemics) är en plattform för datorstödd utbrottsdetektion som är utvecklad på Smittskyddsinstitutet. Systemet är skraddarsytt för bevakningen av de anmälningspliktiga sjukdomarna i Sverige men kan också vara till nytta för andra aktörer som kan använda sig av komponenter i systemet och av arbetssättet. Syftet med den här rapporten är att beskriva hur vi arbetar med systemet och hur det ser ut, för att de som behöver ta fram ett liknande system ska kunna ta del av våra kunskaper och erfarenheter.

Rapporten är i huvudsak en sammanställning och bearbetning av den dokumentation och de vetenskapliga artiklar om systemet som sedan tidigare finns på engelska. Vi sammanfattar dem här för att ge en samlad bild av vårt arbete med datorstödd utbrottsdetektion.

Sammanställningen har gjorts av Anna-Maria Kling, Maria Grünewald och Anette Hulth på Smittskyddsinstitutet. Ett stort tack riktar vi till Myndigheten för samhällsskydd och beredskap, som gett ekonomiskt stöd till projektet. Vi vill också tacka de sjukdomsansvariga handläggarna på Smittskyddsinstitutet och den personal inom hälso- och sjukvården som rapporterar till och arbetar med SmiNet. Sist men inte minst tackar vi de forskare runt om i världen som delat med sig av sina algoritmer och program. Ni har alla bidragit till utvecklingen av systemet genom åren.

Detta kunskapsunderlag är framtaget på Smittskyddsinstitutet. Grundfunktionaliteten i systemet är fortfarande densamma även om vidareutveckling skett. Innehållet i kunskapsunderlaget återger i stora drag hur CASE används på Folkhälsomyndigheten.

Anders Tegnell

Avdelningschef, avdelningen för epidemiologi och utvärdering



# Innehåll

Sammanfattning.....	8
Bakgrund.....	9
Varför datorstödd utbrottsdetektion? .....	9
Datorstödd utbrottsdetektion och kopplingen till SMI:s uppdrag.....	9
Läsinstruktioner.....	10
Datorstödd utbrottsdetektion i praktiken.....	11
Beskrivning av SMI:s system för datorstödd utbrottsdetektion (CASE) .....	11
Hur vi arbetar med CASE.....	12
Utvärdering av användbarheten.....	13
Fördjupad beskrivning av CASE och de statistiska metoderna .....	15
Statistiska metoder .....	15
SaTScan Poisson model.....	16
SaTScan Space Time Permutation model .....	16
Farringtons algoritm .....	16
Outbreak P Statistic .....	17
Tröskelvärdesalgoritmen .....	17
Utvärderingsmodul .....	17
E-postsignaler .....	18
Öppen källkod .....	18
Webbplats .....	18
Diskussion .....	19
Data och statistiska metoder .....	19
Användarnas centrala roll i utvecklingen av CASE .....	20
Framtiden för datorstödd utbrottsdetektion på SMI.....	20
Referenser .....	21

# Sammanfattning

CASE (Computer Assisted Search for Epidemics) är en plattform för datorstödd utbrottsdetektion som är utvecklad på Smittskyddsinstitutet (SMI). Det här dokumentet beskriver hur CASE används och vilka erfarenheter användarna har av systemet. Det beskriver också kortfattat de statistiska metoder som systemet använder för att hitta avvikelser i antalet rapporterade fall av olika smittsamma sjukdomar.

Det främsta syfte med CASE är att varna för möjliga utbrott av smittsamma sjukdomar. Systemet kan också hjälpa sjukdomsansvariga handläggare med mindre erfarenhet av en viss sjukdom att upptäcka avvikelser, något som kan vara användbart exempelvis i semestertider. Målet med systemet är att det ska vara ett värdefullt komplement till det dagliga bevakningsarbetet av de anmälningspliktiga sjukdomarna vid SMI.

Det tekniska systemet finns tillgängligt som öppen källkod. En installationsmanual inkluderas i programvaran och en användarmanual finns att ladda ner. Den beskriver mer i detalj hur de statistiska metoderna fungerar och hur de kan skraddarsys för att passa ändamålet av bevakningen. Publikationer, källkod, manualer och nyheter om CASE finns tillgängliga på <https://case.folkhalsomyndigheten.se/>.

CASE har varit i drift på SMI sedan hösten 2009 och används i dag av flertalet av våra sjukdomsansvariga handläggare. Den första användarstudien av systemet visar att de sjukdomsansvariga handläggarna på SMI upplever CASE som användbart och relevant.



# Bakgrund

## Vad är datorstödd utbrottsdetektion?

Målet med datorstödd utbrottsdetektion är att stödja sjukdomsansvariga handläggare att upptäcka när ett onormalt antal fall (misstänkta eller bekräftade) av en viss sjukdom har rapporterats in. På Smittskyddsinstitutet (SMI) använder vi datorstödd utbrottsdetektion på data om de sjukdomar som är anmälningspliktiga enligt Smittskyddslagen (SFS nr: 2004:168) och som därför kontinuerligt rapporteras in från sjukvården och de mikrobiologiska laboratorerna i Sverige. När man bedömer vad som är onormalt, kan man ta hänsyn till exempelvis geografisk spridning eller smittvägar. Man kan också ta hänsyn till förändringar i antalet fall över tid. Vilka metoder som passar för bevakningen skiljer sig mellan sjukdomarna. Metoderna kan vara allt från enkla tröskelvärden till komplexa statistiska metoder som tar historiska eller geografiska data i beaktande.

Systemet larmar när antalet fall skiljer sig tillräckligt mycket från det förväntade antalet. Sedan avgör en expert, vanligen en epidemiolog, om signalen tyder på ett utbrott och om det krävs vidare utredning. Utredningen syftar till att klarlägga förekomsten, utbredningen, allvarlighetsgraden och orsaken till förändringen i antalet fall. Därefter kan man sätta in lämpliga åtgärder för att begränsa smittspridningen och förebygga nya utbrott av liknande typ. På SMI bevakar vi med datorstödd utbrottsdetektion i dagsläget bara ökning av antalet fall. Men även minskningar kan givetvis vara av intresse, exempelvis om man vill kontrollera att diagnostik och rapportering fungerar som det ska.

## Varför datorstödd utbrottsdetektion?

Det främsta syftet med datorstödd utbrottsdetektion är att varna för möjliga utbrott. Ett sådant system kan göra sjukdomsansvariga handläggare uppmärksamma på avvikelser i ett tidigt skede. Dessutom kan systemet i vissa fall upptäcka utbrott som man annars skulle ha missat. Systemet kan därutöver hjälpa epidemiologer med mindre erfarenhet av en viss sjukdom att upptäcka utbrott. Detta kan vara användbart exempelvis i semestertider eller för ny personal. Systemet kan dock inte ersätta mänskliga experter utan fungerar på SMI som ett komplement och stöd till myndighetens dagliga rutinbevakning; det är just därför vi använder begreppet ”datorstödd” utbrottsdetektion.

Den roll datorstödd utbrottsdetektion spelar i processen att upptäcka utbrott kan verka liten, men tidsaspekten är viktig. Ju tidigare ett utbrott upptäcks, desto tidigare kan myndigheterna utreda källan och sätta in åtgärder för att hindra vidare smittspridning. Målet med datorstödd utbrottsdetektion är alltså att se till att utbrott upptäcks så fort som möjligt.

## Datorstödd utbrottsdetektion och kopplingen till SMI:s uppdrag

Under åren 2006–2011 drev SMI ett projekt med syfte att bygga upp ett datorsystem för utbrottsdetektion för de anmälningspliktiga sjukdomarna. Det var

inom ramen för det här projektet som vi utvecklade systemet Computer Assisted Search for Epidemics (CASE). Projektet finansierades då av Myndigheten för samhällsskydd och beredskap (tidigare Krisberedskapsmyndigheten), men sedan 2012 bekostas det av SMI:s ramanslag.

SMI har ett nationellt perspektiv på bevakningen av smittsamma sjukdomar, medan smittskyddsmyndigheterna runt om i Sverige ansvarar för bevakningen inom sitt landsting. På SMI bevakar ett antal epidemiologer de olika anmälningspliktiga sjukdomarna för att analysera trender och upptäcka utbrott. De planerar också det förebyggande arbetet. På SMI finns även laboratorieverksamhet som är specialiserad på diagnostik, liksom kompetens inom informationssystem och biostatistik. Dessa yrkesgrupper arbetar ofta tillsammans, både i rutinbevakningen av sjukdomar och vid utbrottsutredningar. Den främsta uppgiften för CASE är att hitta utbrott som sträcker sig över mer än ett landsting, speciellt sådana utbrott som bara har få fall i varje landsting. De kan nämligen vara svåra att upptäcka för de regionala och lokala myndigheterna.

## Läsinstruktioner

Söker du en övergripande bild av hur datorstödda system för utbrottsdetektion fungerar och hur det används på SMI? Då rekommenderar vi att du i första hand läser avsnitten *Vad är datorstödd utbrottsdetektion*, *Varför datorstödd utbrottsdetektion* och *Datorstödd utbrottsdetektion och kopplingen till SMI:s uppdrag*. I avsnittet *Datorstödd utbrottsdetektion i praktiken* finns en översiktlig beskrivning av hur CASE används på SMI. I *Fördjupad beskrivning av CASE och de statistiska metoderna* får du en mer detaljerad teknisk beskrivning av systemet liksom en vidare förklaring till de statistiska metoder som används i systemet.

# Datorstödd utbrottsdetektion i praktiken

## Rapportering av anmälningspliktiga sjukdomar till SmiNet

Sverige är indelat i 21 landsting. Varje landsting har en smittskyddsläkare som är ansvarig för bevakningen av de smittsamma sjukdomarna inom sitt landsting och för det förebyggande arbetet. Både smittskyddsläkaren och SMI får en rapport från sjukvården eller från något av Sveriges mikrobiologiska laboratorier när man har ett bekräftat eller misstänkt fall av en anmälningspliktig sjukdom.

Alla de 21 smittskyddsläkarna och majoriteten av sjukhusen och laboratorierna i Sverige har tillgång till en gemensam databas för de anmälningspliktiga sjukdomarna, SmiNet (Rolfhamre m.fl., 2006). Behandlande läkare gör anmälningar via ett webbformulär, medan laboratorierna anmäler verifierade fall direkt. SmiNet innehåller alltså information om både kliniska fall och laborieverifierade prover. År 2011 gjordes 68 619 kliniska anmälningar och 102 333 laboratorieanmälningar (varav 79 % helt automatiskt). I SmiNet kombineras de kliniska anmälningarna med laboratorieresultaten för varje individ och bildar fallrapporter. År 2011 bildades totalt 68 619 fall i databasen.

## Beskrivning av SMI:s system för datorstödd utbrottsdetektion (CASE)

På SMI har vi använt vårt egenutvecklade system CASE sedan augusti 2009 för datorstödd utbrottsbevakning. Det fungerar som ett stöd för våra epidemiologer i den dagliga bevakningen av de anmälningspliktiga sjukdomarna (Cakici m.fl., 2010). CASE hämtar avidentifierade data från SmiNet. Utifrån dessa data gör vi automatiska analyser för att hitta avvikelser i antalet rapporterade fall.

Systemet körs automatiskt en gång per dygn (natttid). Beräkningarna baseras på antalet rapporterade fall, såväl per landsting som nationellt. Om antalet smittade av en viss sjukdom är onormalt högt, skickas automatiskt en varning med e-post till den eller de personer hos oss på SMI som handlägger sjukdomen i fråga. Vad som är ett högt antal avgörs av en eller flera statistiska metoder, som arbetar parallellt och oberoende av varandra.

De epidemiologer som jobbar med bevakning på SMI får regelbundet information om CASE för att avgöra om det kan vara ett bra verktyg i deras arbete. I dagsläget (år 2012) använder de flesta av våra epidemiologer CASE. Epidemiologerna ansvarar för olika sjukdomsgrupper: HIV, andra sexuellt överförbara och blodburna sjukdomar, zoonoser, mat- och vattenburna sjukdomar, sjukdomar som det finns vaccin mot och antibiotikaresistenta smittämnen. Dessa epidemiologer tar emot signaler från CASE för drygt 40 diagnoser och subtyper.

CASE använder för närvarande fem olika statistiska metoder för datorstödd utbrottsdetektion. Metoderna kan användas för att upptäcka utbrott både på diagnosnivå och på subtypsnivå. Olika statistiska metoder kan användas på varje smittämne, och varje metods inställningar kan skräddarsys för varje enskilt smittämne. Valet av statistisk metod beror på epidemiologin för smittämnet och på



Arbetsflödet ser ut på följande sätt när en epidemiolog vill använda CASE för att bevaka en sjukdom: Epidemiologen kontaktar statistikern för ett möte. Vid mötet presenterar statistikern CASE, inklusive syftet med systemet och det grafiska gränssnittet (som bara används av statistikern). Statistikern förklarar även hur de statistiska metoderna fungerar och hur man tolkar informationen i de e-postsignaler som CASE skickar. Dessa steg kan man naturligtvis hoppa över eller gå igenom kort om epidemiologen redan använder CASE för andra sjukdomar. I nästa steg förklarar epidemiologen för statistikern hur sjukdomens epidemiologi ser ut och vilken typ av ökning som bör bevakas. Därefter analyseras historiska data för att få bättre insikt om sjukdomens epidemiologi och därmed vilken statistisk metod som skulle kunna vara lämplig för bevakningen.

CASE-gruppen hos oss på SMI kontaktar regelbundet de epidemiologer som använder systemet. Det gör man för att diskutera hur CASE fungerar för de sjukdomar som bevakas och hur systemet kan vidareutvecklas för att bättre möta epidemiologernas behov. Dessutom samarbetar CASE-gruppen och epidemiologerna för att sätta upp datorstödd utbrottsbevakning för fler smittämnen än de som redan är aktiverade. Statistikern för logg över de diskussioner som förs, både om de signaler som systemet genererat och om förslag på förbättringar av CASE. Alla signaler som systemet skickar till de ansvariga epidemiologerna går även till en särskild e-postlåda för uppföljning och utvärdering. Det är dock epidemiologens uppgift att värdera och agera på de signaler som har skickats ut.

## Utvärdering av användbarheten

CASE-gruppen gör vartannat år en undersökning om hur användarna upplever nyttan av systemet. Den första undersökningen genomfördes i februari 2011 (Kling m.fl., 2012). Alla de nio epidemiologer som då använde CASE bjöds in att besvara en webbenkät. Enkäten bestod av 13 frågor indelade i tre områden: användbarheten av systemet, samarbetet med CASE-gruppen och prioriteringen av den framtida utvecklingen. Frågorna om systemets användbarhet begränsade sig till den tidsperiod som den enskilda epidemiologen hade använt CASE.

Samtliga användare besvarade enkäten (n=9, 100 % svarsfrekvens). I tabell 1 visas svaren på några av de påståenden som de fick ta ställning till i enkäten. Alla användare tyckte att CASE-signalerna bekräftade de signaler de själva såg i det manuella bevakningsarbetet och att signalerna pekade på situationer då fallen i SMI borde granskas mer ingående. Fem av de nio användarna meddelade att CASE hade gett den första signalen om ett riktigt utbrott, det vill säga att CASE till exempel hade varnat innan en smittskyddsläkare kontaktat dem eller innan de själva sett en ökning i SMI. Dessa fem epidemiologer är ansvariga för bevakningen av STI-sjukdomar (n=3) och antibiotikaresistens (n=2).

Alla svarande angav att de känner till vem de ska kontakta om de har frågor om CASE. De tyckte också att frågor om systemet hade besvarats på ett bra sätt. Åtta av nio tyckte att deras idéer och feedback tagits emot och tagits till vara på ett bra sätt av CASE-gruppen (den nionde användaren hade inte gett någon feedback till gruppen). Användarna tyckte att de viktigaste punkterna att prioritera i CASE var att bygga en funktion för att inte få identiska signaler (n=5), att lägga till nya statistiska metoder (n=2) och att generera automatiska analyser på olika smittvägar (n=2).

Tabell 1. Resultat från användarstudien, n=9.

Påstående	Instämmer	Instämmer inte	Vet ej
CASE bekräftar signaler från manuellt bevakningsarbete	9	0	0
CASE ger en indikation om när fallen i SmiNet borde granskas mer ingående	8	0	1
CASE sparar tid i manuellt bevakningsarbete	7	2	0
CASE missar att ge signaler om utbrott	0	5	4
CASE larmar för ofta om icke-relevanta ökningar av antalet fall	3	6	0

# Fördjupad beskrivning av CASE och de statistiska metoderna

## Systembeskrivning

CASE består av tre komponenter som sköter *konfigurering*, *extrahering* av data och *detektion* av avvikelser i data. Konfigurationsdelen innehåller det grafiska gränssnittet där administratören anger vilka statistiska metoder och parametrar som ska användas för de olika smittämnen och vilka epidemiologer som ska ta emot signaler från systemet för respektive smittämne. Extraktionsdelen används för att kopiera data från SmiNet till en lokal databas. Denna databas lagrar även information om de inställningar som görs i gränssnittet. Detektionsdelen körs automatiskt en gång per dygn och tillämpar då de statistiska metoder som valts på alla aktiverade smittämnen. När systemet skapar signaler, skickar det ut larmen som e-postmeddelanden. Om flera statistiska metoder är aktiverad för ett smittämne och dessa larmar samtidigt, kommer separata e-postmeddelanden att skickas ut för de olika metoderna. Samtidigt skapar systemet automatiskt detaljerade loggar över hela processen.

CASE bygger sina analyser på information från SmiNet. SmiNet skapar en fallrapport när systemet tar emot en klinisk rapport eller en laborierapport för någon anmälningspliktig sjukdom. Men fallrapporter skapas som regel bara om det sedan tidigare inte finns en fallrapport för just denna sjukdomsepisod hos en och samma patient. Skulle flera rapporter komma in för samma fall, läggs den kompletterande informationen till den redan existerande rapporten. En patient kan dock ibland få flera fallrapporter, trots att diagnosen är densamma. Det sker om det har det gått lång tid mellan två rapporter för just den patienten, eftersom det då är troligt att det rör sig om upprepade smittor snarare än flera rapporter om samma sjukdomsepisod. Smittämne, datum, landsting, smittväg och infektionsland kopieras sedan till CASE:s databas för varje fall. Men CASE använder ingen information som kan kopplas till en enskild patients identitet.

När data kopieras från SmiNet till CASE:s lokala databas, används fältet *statistikdatum* för att datera fallet. Detta datum sätts automatiskt när ett fall med en viss diagnos dyker upp i SmiNet för första gången, och det är det enda datum som finns på samtliga rapporter i SmiNet. För varje fallrapport i SmiNet finns det utöver detta ungefär tjugo andra olika datum, men det varierar från fall till fall vilka som är ifyllda. Vissa av dessa datum skapas automatiskt av systemet, medan andra läggs in av sjukvården eller laboratoriet som rapporterar fallet. Ett datum som bättre skulle spegla när patienten blev sjuk är provtagningsdatumet. Provtagningsdatumet saknas dock i en stor del av fallrapporterna.

## Statistiska metoder

CASE använder i dagsläget fem statistiska metoder (även kallat *algoritmer*). Fyra av dessa är standardalgoritmer på området, medan den femte är speciellt framtagen inom CASE-projektet. Den enklaste algoritmen i CASE är en tröskelvärdesalgoritm, där larmgränsen sätts manuellt beroende på smittämne. Två

mer komplexa algoritmer kommer från programvaran SaTScan (2011), som kan laddas ner gratis från internet och som är utvecklad för analys av data som varierar över tid och geografiska områden. Dessa båda algoritmer tar hänsyn till befolkningsstrukturen. En fjärde algoritm (Farrington m.fl., 1996) analyserar fall på nationell nivå utifrån historiska data. Den kommer från programpaketet *Surveillance* (Höhle, 2007) i den statistiska programvaran R (R Development Core Team, 2012). Den femte algoritmen undersöker om antalet fall ökar mer än förväntat (Frisén, Andersson & Schiöler, 2009). Den programvaran kan laddas ner gratis från internet (Göteborgs universitet, 2011). CASE är också förberett för fler algoritmer om ett sådant behov skulle dyka upp i framtiden.

Nedan beskriver vi de olika algoritmerna mer i detalj. Vi ger dessutom exempel på hur man kan avgöra när en algoritm passar att använda.

### SaTScan Poisson model

*SaTScan discrete Poisson model* (Kulldorff, 1997) från programvaran SaTScan används för att undersöka geografiska anhopningar av fall under en specifik period genom att använda en så kallad Poisson-modell. Algoritmen bevakar geografiska anhopningar, som kan bestå av ett eller flera intilliggande områden, i vårt fall landsting. Den tar även hänsyn till populationsstorleken i de olika landstingen. Algoritmen signalerar sedan när antalet fall inom ett geografiskt område statistiskt signifikant skiljer sig från vad som är förväntat.

Algoritmen är exempelvis lämplig att använda när:

- fall- och befolkningsdata är tillgängliga på landstingsnivå
- antalet nya fall (incidensen) är på liknande nivåer inom varje landsting
- utbrott förväntas bestå av geografiska anhopningar av fall

### SaTScan Space Time Permutation model

*SaTScan Space Time Permutation model* (Kulldorff m.fl., 2005) undersöker förändringar av fall över tid består av geografiska anhopningar. Om incidensen förändras överallt, eller om ett område har en ständigt hög incidens, ser algoritmen inget utbrott. Algoritmen tar också hänsyn till exempelvis säsongsvariation och demografisk spridning. Om däremot incidensen ökar i något område utan att man ser motsvarande förändring i resten av landet, tolkar algoritmen det som ett utbrott. Algoritmen tar också hänsyn till var i landet de olika områdena ligger i förhållande till varandra.

Algoritmen är exempelvis lämplig att använda när:

- fall- och befolkningsdata är tillgängliga på landstingsnivå
- utbrott förväntas bestå av geografiska anhopningar av fall

### Farringtons algoritm

Denna algoritm är utvecklad och används av Health Protection Agency i England och Wales i Storbritannien (Farrington m.fl., 1996). Algoritmen, som vi i CASE



kallar Farrington-algoritmen, använder aggregerade falldata på nationell nivå för att undersöka om antalet fall av ett smittämne överstiger den förväntade nivån. Den förväntade nivån sätts utifrån tidigare års data, och algoritmen ger signal när antalet fall överstiger detta tröskelvärde. En vecka för Farrington-algoritmen börjar på en måndag och slutar på en söndag, och antalet fall adderas över veckan. Eftersom CASE körs varje natt, är en signal alltså mer trolig i slutet på veckan.

Algoritmen är exempelvis lämplig att använda när:

- antalet fall varierar mellan år på regelbundet sätt
- historiska data finns tillgängliga

### Outbreak P Statistic

Algoritmen *Outbreak P statistic* (Frisén, Andersson & Schiöler, 2009) används i CASE för att undersöka om antalet fall per vecka ökar mer än förväntat på nationell nivå. Algoritmen ger en signal om värdet (statistikan) är högre än ett förutbestämt gränsvärde. För att mäta ökningen tar statistikan hänsyn både till antalet fall innevarande vecka och hur det sett ut tidigare veckor.

Algoritmen är exempelvis lämplig att använda när:

- fördelningen av ett smittämne är stabil över tid
- det inte finns någon säsongsvariation.

### Tröskelvärdesalgoritmen

Tröskelvärdesalgoritmen är framtagen av CASE-gruppen. Den ger signal när antalet fall på nationell nivå överstiger en tröskelgräns under en viss tidsperiod. Tidsperioden justeras beroende på smittämne och tröskelgränsen sätts manuellt.

Algoritmen är exempelvis lämplig att använda när:

- en signal ska genereras vid en specifik tröskelgräns för en förutbestämd tidsperiod
- smittämnet är ovanligt, då sätts tröskelvärdet till noll (0)
- en sjukdom inte varierar med säsong.

## Utvärderingsmodul

Det är en utmaning att avgöra vilka algoritmer och parameterinställningar man ska tillämpa när man använder datorstödd utbrottsdetektion. Det är viktigt att inte alltför många utbrott missas. Samtidigt får antalet falska varningar om utbrott inte bli oacceptabelt högt. För att underlätta detta arbete har vi vidareutvecklat CASE med en så kallad utvärderingsmodul. I utvärderingsmodulen kan man köra algoritmerna på riktiga data från ett valfritt antal dagar bakåt i tiden. På så vis får man en mer samlad bild av hur bra algoritmerna fungerar för den sjukdom man är intresserad av. Resultaten presenteras både i en grafisk sammanställning och som en datafil som kan användas för vidare analyser i externa program.

Utvärderingsmodulen minskar mängden ”trial and error” som behövs för att välja

algoritmer och parameterinställningar. Detta gör resultaten mer pålitliga och ökar användbarheten hos systemet.

## E-postsignaler

CASE genererar signaler om möjliga utbrott och skickar dem till den ansvariga epidemiologen via e-post. Eftersom CASE gör sina automatiska analyser efter midnatt varje dag, finns dagsfärska CASE-signaler i epidemiologernas inkorgar varje morgon.

Alla e-postsignaler från CASE innehåller information om smittämne, vilken statistisk metod som larmat, antal fall och så vidare. Dessutom innehåller de kort, generell information om hur signalen kan tolkas.

## Öppen källkod

CASE finns tillgängligt som öppen källkod under GNU General Public License version 3 (Free Software Foundation, 2007) och kan laddas ner från CASE:s webbplats. Användargränssnittet och det mesta av dokumentationen är på engelska.

CASE är utvecklat i Java för både Windows och Linux operativsystem. För att köra CASE behövs en server med Sun Java Runtime Environment 6.0 (eller högre). Det finns också andra mjukvaror som måste installeras separat och som varierar beroende på operativsystem. Dessa mjukvaror är följande:

### Windows

- R för Windows [<http://ftp.sunet.se/pub/lang/CRAN/>]

### Linux

- R för Linux [<http://ftp.sunet.se/pub/lang/CRAN/>]
- SaTScan för Linux  
[[http://www.satscan.org/download\\_satscan\\_for\\_linux.html](http://www.satscan.org/download_satscan_for_linux.html)]
- ImageMagick för Unix [<http://www.imagemagick.org/script/index.php>]

Alla script för att skapa och konfigurera den lokala databasen är inkluderade i CASE-mjukvaran.

## Webbplats

CASE har en webbplats där publikationer, källkod, manualer och nyheter om CASE finns tillgängliga: <https://smisvn.smi.se/case/>.

# Diskussion

Arbetet med datorstödd utbrottsdetektion på SMI har vuxit från idé till ett egenutvecklat datorprogram, CASE. Under implementeringsfasen blev det tydligt att framför allt samarbetet med användarna var viktigt för att skapa ett väl fungerande system. Programmet används nu på SMI i det dagliga bevakningsarbetet av de anmälningspliktiga sjukdomarna.

## Data och statistiska metoder

CASE gör det möjligt för oss att automatiskt analysera falldata för hundratals olika smittämnen för olika delar av Sveriges befolkning på kort tid. De olika metoderna i CASE, allt från den enkla tröskelvärdesalgoritmen till de komplexa statistiska metoderna, och att man kan göra smittämnes-specifika parameterinställningar betyder att systemet kan anpassas till datorstödd bevakning av många olika typer av smittämnen.

Det händer dock att man får falska positiva signaler när man använder statistiska metoder för bevakning. Man kan ta bort en del av dessa felsignaler genom att finjustera inställningarna för metoderna, men på grund av slumpvariation i datamängden kommer metoderna alltid att ge en del signaler även när det inte är sanna utbrott. Med CASE bevakar vi många sjukdomar med flera olika metoder samtidigt, vilket ökar antalet falska positiva signaler. Hur känsliga inställningar en användare vill ha är en individuell bedömning, där man väger in både hur viktigt det är att fånga upp verkliga signaler och hur mycket arbetstid det tar i anspråk att hantera signaler där det inte handlar om riktiga utbrott.

Enkla metoder som appliceras på data av hög kvalitet producerar ofta bättre resultat än komplexa metoder på data av låg kvalitet. Därför är kvaliteten på de data som ligger till grund till bevakningen den viktigaste faktorn när man skapar ett system för datorstödd sjukdomsbevakning. Om de data som ska användas inte är pålitliga, är det bättre att lägga resurser på att förbättra kvaliteten på datainsamlingen än att bygga upp ett datoriserat bevakningssystem.

Hittills har kvaliteten på signalerna från CASE utvärderats genom kommunikation med epidemiologerna. Detta är inte optimalt av olika anledningar, och det är något vi skulle vilja förbättra. I framtiden kan nya data från SMI bli en värdefull tillgång för att kunna utvärdera CASE-signalerna på ett mer systematiskt sätt. De som utvecklar SMI håller nämligen nu på att ta fram en ny modul, som kan lagra information om vilka fallrapporter som anses tillhöra ett visst utbrott. Tidigare har sådan information inte registrerats i SMI.

Här vill vi också passa på att lyfta fram hur viktigt det är att man har realistiska förväntningarna på ett system som CASE. Ett datoriserat bevakningssystem kan upptäcka pågående utbrott och till exempel ökningarna som beror på säsongseffekter. Däremot kan vi inte förutspå framtida utbrott utifrån de fallrapporter som finns i SMI.

## Användarnas centrala roll i utvecklingen av CASE

Användarstudien från år 2011 visar att CASE är ett viktigt och uppskattat verktyg i det dagliga bevakningsarbetet av de anmälningspliktiga sjukdomarna. Synpunkter från användare har varit ovärderliga för att skapa ett användbart system, allt sedan uppbyggnaden av datorstödd utbrottsdetektion startade på SMI. Användarnas roll inom CASE-arbetet är fortfarande mycket central, vilket görs tydligt i rutinbeskrivningen för CASE. CASE-gruppen har också en pågående dialog med användarna och samlar bland annat in deras åsikter och förbättringssynpunkter genom en användarundersökning vartannat år.

Under den tid som CASE har använts på SMI, har systemet kontinuerligt anpassats efter användarnas önskemål. Till exempel kan epidemiologerna nu välja om systemet ska analysera alla rapporterade fall eller om det endast ska inkludera inhemskt smittade. För Hepatit B har man gjort en speciallösning, där endast akuta fall inkluderas i analyserna. Numera kan epidemiologerna också själva välja att avstå från upprepade utskick av en signal som redan har skickats. Det var ett av de önskemål som fördes fram i användarundersökningen 2011. Just nu pågår en vidareutveckling av CASE som kommer att göra det möjligt att bevaka smittämnen uppdelat på smittvägar.

## Framtiden för datorstödd utbrottsdetektion på SMI

Vidareutvecklingen av CASE kommer att fokusera på att öka kvaliteten på de signaler som CASE skickar. Det kan bland annat ske genom att vi ökar kvaliteten på de data som CASE använder. Detta är ett kontinuerligt pågående arbete på SMI som långsiktigt kan påverka systemets funktionalitet. Ett annat sätt att mer direkt öka kvaliteten på signalerna är att se till att systemet tillhandahåller statistiska metoder som är lämpliga för den bevakning som epidemiologerna önskar. Vi kan också förbättra den information som följer med signalerna, så att epidemiologerna lättare kan avgöra hur relevanta de är.

Flera nationella institut i Europa använder system för datorstödd utbrottsdetektion (Hulth m.fl., 2010). Av dessa system är CASE det system som erbjuder mest anpassningsmöjligheter till olika smittämnen och användare. Just flexibiliteten i CASE tror vi är en av flera framgångsfaktorer för systemet, det vill säga att valet av algoritm och parameterinställningar kan vara olika för olika smittämnen och att systemet kan skräddarsys för olika behov. Andra faktorer är det nära samarbetet mellan CASE-gruppen och epidemiologerna, tydliga roller och rutinbeskrivningar inom CASE-gruppen, liksom den vidareutveckling som gör att systemet kontinuerligt anpassas efter epidemiologernas önskemål.

Vi kommer att göra en andra användarundersökning under våren 2013. Resultaten kommer att vara styrande i vårt vidare arbete med datorstödd utbrottsdetektion, så att CASE ska fortsätta vara ett användbart verktyg i den dagliga rutinbevakningen av smittsamma sjukdomar på SMI.

# Referenser

- Cakici B, Hebing K, Grünewald M, Saretok P, Hulth A. 2010. CASE: a framework for computer supported outbreak detection. *BMC Med Inform Decis Mak.* 10:14.
- Farrington CP, Andrews NJ, Beale AD, Catchpole MA. 1996. A statistical algorithm for the early detection of outbreaks of infectious disease. *J Roy Stat Soc.* 159(3):547–563.
- Free Software Foundation. 2007. Gnu General Public License Version 3, 29 June 2007. <http://www.gnu.org/licenses/gpl.html>. Hämtad 2012-10-15.
- Frisén M, Andersson E, Schiöler L. 2009. Robust outbreak surveillance of epidemics in Sweden. *Stat Med.* 28(3):476–493.
- Göteborgs universitet. 2011. Outbreak detection. [http://www.economics.handels.gu.se/english/Units+and+Centra/statistical\\_research\\_unit/software/](http://www.economics.handels.gu.se/english/Units+and+Centra/statistical_research_unit/software/). Hämtad 2012-10-15.
- Hulth A, Andrews N, Ethelberg S, Dreesman J, Faensen D, van Pelt W, Schnitzler J. 2010. Practical usage of computer-supported outbreak detection in five European countries. *Euro Surveill.* 15(36).
- Höhle M. 2007. Surveillance: An R package for the monitoring of infectious diseases. *Comput Stat.* 22(4):571–582.
- Kling A-M, Hebing K, Grünewald M, Hulth A. 2012. Two years of computer supported outbreak detection in Sweden: the user's perspective. *J Health Med Informat.* 3:108.
- Kulldorff M. 1997. A spatial scan statistic. *Commun Stat Theory Methods* 26(6):1481–1496.
- Kulldorff M, Heffernan R, Hartman J, Assunção RM, Mostashari F. 2005. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med.* 2(3):e59.
- R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/> Hämtad 2012-10-15.
- Rolfhamre P, Jansson A, Arneborn M, Ekdahl K. 2006. SmiNet-2: Description of an internet-based surveillance system for communicable diseases in Sweden. *Euro Surveill.* 11(5):103–7.
- SaTScan. 2011. Software for the spatial, temporal, and space-time scan statistics. <http://www.satscan.org>. Hämtad 2012-10-15

CASE (Computer Assisted Search for Epidemics) är en plattform för datorstödd utbrottsdetektion som är utvecklad på Smittskyddsinstitutet (SMI). Det här dokumentet beskriver hur CASE används på myndigheten och vilka erfarenheter användarna har av systemet. Det beskriver också kortfattat de statistiska metoder som systemet använder för att hitta avvikelser i antalet rapporterade fall av olika smittsamma sjukdomar. Underlaget riktar sig till aktörer som vill veta mer om hur myndigheten arbetar med datorstödd utbrottsdetektion.

-----  
*Folkhälsomyndigheten är en nationell kunskapsmyndighet som arbetar för en bättre folkhälsa. Det gör myndigheten genom att utveckla och stödja samhällets arbete med att främja hälsa, förebygga ohälsa och skydda mot hälsorhot.*

*Vår vision är en folkhälsa som stärker samhällets utveckling.*



Folkhälsomyndigheten

**Solna** Nobels väg 18, SE-171 82 Solna **Östersund** Forskarens väg 3, SE-831 40 Östersund.

[www.folkhalsomyndigheten.se](http://www.folkhalsomyndigheten.se)