

ارزشیابی عملکرد الگوریتم مجموع تراکمی در کشف طغیان

منوچهر کرمی^۱، حمید سوری^۲، یدالله محرابی^۳، علی اکبر حقدوست^۴، محمد مهدی گویا^۵، نادر اسماعیل نسب^۶

^۱ استادیار اپیدمیولوژی، مرکز تحقیقات مدل‌سازی بیماری‌های غیرواگیر و گروه آمارزیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، ایران

^۲ استاد اپیدمیولوژی، مرکز تحقیقات ارتقاء ایمنی و پیشگیری از مصدومیت‌ها و گروه اپیدمیولوژی، دانشگاه علوم پزشکی شهید بهشتی، ایران

^۳ استاد آمار زیستی، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی، ایران

^۴ استادیار اپیدمیولوژی، مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، ایران

^۵ استاد بیماری‌های عفونی، مرکز مدیریت بیماری‌های واگیر، وزارت بهداشت، درمان و آموزش پزشکی، ایران

^۶ دانشیار اپیدمیولوژی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت کردستان، گروه اپیدمیولوژی و آمار، دانشکده پزشکی، دانشگاه علوم پزشکی کردستان، سنندج، ایران

نویسنده مسئول: منوچهر کرمی، نشانی: دانشگاه علوم پزشکی همدان، دانشکده بهداشت، گروه آمارزیستی و اپیدمیولوژی. کد پستی: ۶۵۱۷۸۳۸۷۳۶، پست الکترونیک:

ma.karami@umsha.ac.ir

تاریخ دریافت: ۹۱/۴/۶؛ پذیرش: ۹۱/۱۱/۷

مقدمه و اهداف: کشف زودرس و پاسخ به‌هنگام به بیماری‌های نوپدید و طغیان‌ها یکی از مهم‌ترین اولویت‌های بهداشت عمومی و نظام‌های سلامت است. مطالعه‌هایی بسیار محدود در دنیا به‌طور هم‌زمان برای ارزشیابی روش‌های کشف طغیان، روی داده‌های یکسان و با استفاده از رویکرد ارزشیابی عملکرد بر مبنای داده‌های نیمه‌شبیه‌سازی شده، اقدام کرده‌اند. این مطالعه نیز با هدف ارزشیابی عملکرد الگوریتم مجموع تراکمی در شناسایی به‌هنگام ۸۳۱ روز طغیان نیمه‌شبیه‌سازی شده، انجام شده است.

روش کار: عملکرد الگوریتم مجموع تراکمی در کشف طغیان‌های نیمه‌شبیه‌سازی شده با استفاده از داده‌های موارد گزارش شده روزانه مشکوک به سرخک، طی سال‌های ۱۳۸۷ تا ۱۳۸۹، ارزشیابی گردید. عملکرد این الگوریتم با استفاده از شاخص‌های حساسیت، درصد هشدار کاذب، نسبت‌های درست‌نمایی و سطح زیر منحنی راک ارزیابی شد.

نتایج: حساسیت الگوریتم مجموع تراکمی در شناسایی طغیان‌های شبیه‌سازی شده ۵۰ درصد، با فاصله اطمینان ۹۵ درصد (۵۴-۴۷) بود. عملکرد این الگوریتم بسته به اندازه، شکل و دوره طغیان شبیه‌سازی شده متفاوت است. از نظر شاخص به‌هنگام‌بودن نیز، به‌طور متوسط، کمترین فاصله زمانی بین روز واقعی شروع طغیان بر مبنای داده‌های نیمه‌شبیه‌سازی شده تا زمان کشف طغیان، با این الگوریتم، ۱۳/۸۴ روز بود.

نتیجه‌گیری: اگرچه نتایج مطالعه بیانگر عملکرد خوب الگوریتم مجموع تراکمی در شناسایی طغیان‌های با اندازه بزرگ و دوره کوتاه بود، اما عملکرد این الگوریتم، به‌ویژه در شناسایی طغیان‌های با دوره طولانی و اندازه کوچک، ضعیف بود.

واژگان کلیدی: نظام مراقبت، سرخک، روش کشف طغیان، الگوریتم مجموع تراکمی، روش نیمه‌شبیه‌سازی، ایران

مقدمه

نظام‌های مراقبت سندرمیک را نیز می‌توان در دو گروه منابع داده‌های بالینی، مانند سندروم‌ها و منابع غیر بالینی، از جمله غیبت از کار و مدرسه، دسته‌بندی کرد. با این وجود، عملکرد بهینه نظام‌های مراقبت سندرمیک در گروه استفاده هم‌زمان از بیش از یک یا دو منبع است (۴).

ابزار نظام‌های مراقبت سندرومیک برای پایش تغییرات در روند بیماری‌ها، الگوریتم‌ها یا روش‌های کشف طغیانند که در دو گروه روش‌های زمانی و مکانی قرار می‌گیرند. شناخته‌شده‌ترین روش، استفاده از الگوریتم‌های مبتنی بر کنترل فرآیندهای آماری، مانند الگوریتم مجموع تراکمی^۲ است (۵). برای نمونه، نظام مراقبت

آمدگی، کشف زودرس و پاسخ به‌هنگام به بیماری‌های نوپدید و طغیان‌ها یکی از مهم‌ترین اولویت‌های بهداشت عمومی و نظام‌های سلامت است. با توجه به محدودیت نظام‌های مراقبت رایج در این رابطه، نظام مراقبت سندرمیک که مهم‌ترین ویژگی آن شناسایی زودهنگام طغیان‌ها است، شکل گرفته است (۲-۱). مرکز پیشگیری و کنترل بیماری‌های آمریکا^۱ نظام مراقبت سندرمیک را رویکردی کاوش‌گرایانه توصیف کرده که برای پایش علائم، نشانه‌ها و یا دیگر شاخص‌های بیماری در زمان وقوع رخداد یا نزدیک به آن، برای شناسایی به‌هنگام و سریع‌تر، طغیان نسبت به نظام مراقبت رایج، طراحی شده است (۳). منابع مورد استفاده در

^۲ Cumulative Sums (CUSUM)

^۱ The center for diseases control and prevention(CDC)

مراقبت بهره گرفته‌اند و یا آنکه در ارزشیابی عملکرد و مقایسه روش‌های کشف طغیان از شاخص‌های مختلف، به‌طور همزمان، استفاده کرده‌اند. افزون بر این، سرخک یک بیماری بثوری تبار و بسیار واگیر است که با وجود داشتن واکسن ایمن و اثربخش، همچنان جزو علت‌های عمده مرگ کودکان در سراسر دنیا به‌شمار می‌رود (۱۲). از این رو، برقراری نظام مراقبت مناسب برای بیماری سرخک یکی از مهم‌ترین راهکارهای سازمان جهانی بهداشت برای حذف و کاهش مرگ و میر ناشی از این بیماری است (۱۳). به همین دلیل، مطالعه حاضر با هدف مقایسه عملکرد الگوریتم کشف طغیان مجموع تراکمی با استفاده از رویکرد نیمه‌شبیه‌سازی، بر مبنای داده‌های واقعی موارد گزارش‌شده مظنون به سرخک، طی سال‌های ۱۳۸۷ تا ۱۳۸۹ در ایران، با در نظر گرفتن پارامترهای شکل، اندازه و دوره طغیان‌های رخ داده اخیر و توجه به دانش پویایی انتقال بیماری طراحی شده‌است.

روش کار

در این مطالعه عملکرد الگوریتم کشف طغیان مجموع تراکمی، روی طغیان‌های شبیه‌سازی شده، بر مبنای داده‌های واقعی به شرح زیر بررسی و ارزشیابی شده‌است.

منبع داده مورد استفاده برای آزمون روش‌های کشف طغیان

الگوریتم کشف طغیان مجموع تراکمی روی داده‌های موارد گزارش شده روزانه مشکوک به سرخک، آزموده شده است. با توجه به اهداف مطالعه حاضر، داده‌های کشوری موارد مشکوک به سرخک طی سال‌های ۱۳۸۷ تا ۱۳۸۹، برای پیش‌پردازش و افزودن به طغیان‌های استخراج شده، به شرح زیر به کار رفتند. این داده‌ها شامل ۴۸۳۴ مورد از افراد دارای علائم بالینی تب، بثورات جلدی ماکولوپاپولر و یکی از علائم سرفه یا آبریزش بینی یا قرمزی ملتحمه‌اند که به‌صورت روزانه، از سوی سطوح محیطی نظام مراقبت به مرکز مدیریت بیماری‌های واگیر وزارت بهداشت گزارش شده‌اند.

شبیه‌سازی طغیان‌ها بر مبنای داده‌های واقعی

با توجه به روش‌شناسی رویکرد مورد استفاده برای ارزشیابی عملکرد روش‌های کشف طغیان، یعنی رویکرد نیمه‌شبیه‌سازی، ابتدا، نویسندگان منحنی‌های همه‌گیری بیماری سرخک را از سه منبع متفاوت استخراج کردند و سپس، این طغیان‌ها به داده‌های

BioSense (CDC) با بهره‌گیری از الگوریتم‌های مجموع تراکمی و میانگین متحرک وزن داده شده نمای^۱، یازده سنندرم را تحت مراقبت دارد و دانشگاه جان‌هاپکینز نظام مراقبت ESSENCE را برای مراقبت هشت سنندرم، با استفاده از الگوریتم‌های یادشده و دو روش دیگر راه‌اندازی کرده است (۷-۶). آگاهی از عملکرد نظام‌های مراقبت سنندرمیک در شرایط متفاوت، بسیار مهم است، به طوری که اطلاع از میزان خطای هر نظام در تصمیم‌گیری درباره روش بررسی هشدار کمک می‌کند. از این رو، ارزیابی کامل الگوریتم‌ها و روش‌های کشف طغیان، امکان دستیابی به نقاط قوت و ضعف و نیز کاربرد آن‌ها در شرایط مختلف را فراهم می‌کند. نتایج مطالعه‌هایی که به ارزشیابی این الگوریتم‌ها پرداخته‌اند، بر این نکته توافق دارند که هیچ الگوریتم خاصی به‌تنهایی نمی‌تواند در همه موارد، بهترین عملکرد را داشته باشد (۹-۸).

روش‌های ارزشیابی عملکرد الگوریتم‌های یاد شده در نظام مراقبت سنندرمیک، شامل سه روش ارزشیابی عملکرد بر مبنای داده‌های واقعی^۲ حاصل از نظام‌های مراقبت سنندرمیک، ارزشیابی عملکرد بر مبنای داده‌های نیمه شبیه‌سازی^۳ شده و کاملاً شبیه‌سازی^۴ شده‌اند. روش نیمه‌شبیه‌سازی شده، به دلیل انعطاف‌پذیری و امکان بررسی کامل الگوریتم در شرایط مختلف و نیز، داشتن درجه‌ای بالاتر از روایی، نسبت به روش کاملاً شبیه‌سازی شده، بیشتر اهمیت دارد (۱۱-۱۰، ۵). مهم‌ترین شاخص‌های ارزشیابی عملکرد الگوریتم‌های کشف طغیان نیز شامل حساسیت، ویژگی، درصد هشدارهای کاذب، به‌هنگام‌بودن و سطح زیر منحنی راک^۵ است (۱۱، ۱).

با توجه به فراهم شدن زمینه لازم برای راه‌اندازی نظام مراقبت سنندرمیک در ایران از اواخر سال ۱۳۸۹ و تلاش‌های صورت‌گرفته در این راه، بنا به نیاز مبرم کشور و تعهدات وزارت بهداشت به قوانین بین‌المللی بهداشتی^۶، به نظر می‌رسد نتایج مطالعه حاضر، که بر اساس دانش گروه پژوهش تاکنون مشابه آن در کشور انجام نشده است، در مسیر راه‌اندازی، پایلوت و ارزشیابی عملکرد نظام یاد شده در کشور بسیار یاری‌گر خواهد بود. مطالعه‌هایی بسیار محدود در دنیا، از رویکرد نیمه‌شبیه‌سازی شده، با در نظر گرفتن ایجاد طغیان‌های تقریباً واقعی و دانش کنونی بیماری تحت

^۱ The Exponentially Weighted Moving Average (EWMA)

^۲ Real data testing

^۳ Semi-synthetic simulation

^۴ Fully synthetic simulation

^۵ Area under the Receiver Operating Characteristic (ROC) curve

^۶ International Health Regulations

نیمه‌شبیه‌سازی شده، ارزشیابی گردید. این الگوریتم در شناسایی تغییرات کوچک و متوالی در داده‌ها بسیار مفید است. در واقع، الگوریتم مجموع تراکمی، تجمیع انحراف از مقدار مورد نظر را پیش می‌کند. در این روش یک آماره روزانه از تعداد موارد مظنون به سرخک با استفاده از فرمول زیر برآورد شد (۲، ۲۴):

$$CUSUM_t = \text{Max}(0, CUSUM_{t-1} + Y_t - \sigma/2)$$

که در آن

$$Y_t \text{ تعداد مشاهده‌ها در روز } t \text{ است } (t = 1, 2, \dots, n)$$

σ انحراف معیار مقادیر مشاهده شده است

$CUSUM_{t-1}$ تعداد برآورد شده در زمان $t-1$ ، یعنی آماره روز پیش است.

حد بالای کنترل، به‌عنوان سطح آستانه با استفاده از معادله ۲ برآورد می‌شود و در صورتی که مقدار $CUSUM_t$ بزرگ‌تر از آن باشد، هشدار طغیان اعلان می‌گردد.

$$\text{Upper Control Limit} = UCL = \mu + h \times \sigma \text{ (معادله ۲)}$$

که در آن μ میانگین، σ انحراف معیار مشاهده‌ها و h عددی ثابت است که در این مطالعه برابر با $2/5$ تعیین شده است. در مجموع، ۱۲ حالت مختلف از الگوریتم‌های مجموع تراکمی که به نام‌های $CUSUM_1$ تا $CUSUM_{12}$ در قسمت نتایج نام‌گذاری شده‌اند، با در نظر گرفتن سطوح آستانه و نیز دوره‌های زمانی متفاوت برای برآورد میانگین مشاهده‌ها (در روزهای بدون طغیان) بررسی شدند. دوره‌های زمانی متفاوت برای آزمون الگوریتم‌ها در این مطالعه، از متوسط موارد گزارش شده روزانه مشکوک به سرخک در شرایط معمول و رخ‌ندادن همه‌گیری طی سال‌های ۱۳۸۵ و ۱۳۸۶ در کشور، برآورد شده است.

پیش‌پردازش داده‌ها: رویکردهای شناسایی و حذف الگوهای توجیه‌پذیر

برای پرهیز از پیامدهای آزمون الگوریتم‌های کشف طغیان روی داده‌های خام، از جمله ایجاد هشدارهای کاذب به دلیل مصداق نداشتن پیش‌فرض‌های مربوط به استفاده از این الگوریتم‌ها و وجود احتمالی الگوهای توجیه‌پذیر، پیش از آزمون روش‌های کشف طغیان، الگوهای یاد شده شناسایی و حذف شدند که این فرایند پیش‌پردازش نام دارد. به‌طوری که برای آزمون الگوریتم مجموع تراکمی از مقادیر پیش‌پردازش شده به‌جای داده‌های خام موارد مشکوک به سرخک استفاده شد. رویکردهای شناسایی و حذف الگوهای توجیه‌پذیر با اشاره به روش‌شناسی آن در جای دیگر آورده شده است (۲۶-۲۵).

پیش‌پردازش شده موارد مشکوک به سرخک طی سال‌های ۱۳۸۷ تا ۱۳۸۹ افزوده شد. فواصل افزودن، به‌طور مشخص، یک هفته پس از پایان دوره همه‌گیری پیشین تعیین شد. منحنی‌های همه‌گیری با هدف قابلیت تعمیم به شرایط موجود و آتی رویدادهای احتمالی طغیان سرخک در کشور ایران از سه منبع متفاوت زیر به‌دست آمد:

۱- استخراج منحنی همه‌گیری‌های واقعی بیماری سرخک از منابع موجود: منحنی‌های مربوط به ۱۰ همه‌گیری بیماری سرخک با استفاده از جست‌وجو در منابع موجود (۲۳-۱۴)، به‌شرح منابع شماره ۱۴ تا ۲۳ که داده‌های آن‌ها روزانه گزارش شده بود، برای ارزشیابی عملکرد الگوریتم‌های کشف طغیان، به داده‌های پیش‌پردازش شده موارد گزارش شده سرخک اضافه شدند.

۲- استخراج منحنی همه‌گیری‌های واقعی بیماری سرخک از داده‌های نظام مراقبت این بیماری:

منحنی همه‌گیری مربوط به طغیان‌های واقعی سرخک در کشور در سال ۱۳۸۹ شامل دو مورد طغیان در بندرعباس، یک مورد در جاسک و یک مورد در مشهد که داده‌های آن‌ها در دسترس است، استخراج شد.

۳- شبیه‌سازی ۹ مورد منحنی همه‌گیری با اندازه، شکل و دوره طغیان متفاوت:

برای اطمینان از پوشش بهتر طغیان‌های احتمالی سرخک و ارزشیابی عملکرد الگوریتم‌های مورد آزمون در این مطالعه، منحنی همه‌گیری ۳ مورد طغیان با دوره ۷ روزه، ۳ مورد با دوره ۱۴ روزه و ۳ مورد با دوره ۲۱ روزه شبیه‌سازی شد. هریک از طغیان‌ها به‌دلیل داشتن توزیع‌های متفاوت، شامل ازدیاد خطی، ازدیاد نمایی یا توزیع یکنواخت، از نظر اندازه و شکل متفاوت از یکدیگرند. گفتنی است در طغیان‌های یاد شده حداقل یک مورد و حداکثر ۱۴ مورد (به اندازه ۴ برابر انحراف معیار موارد مشکوک به سرخک دیده شده) به داده‌های پیش‌پردازش شده موارد مشکوک به سرخک افزوده شده است. در مجموع با استفاده از سه منبع یاد شده، ۸۳۱ روز طغیان با هدف ارزشیابی عملکرد الگوریتم‌های متفاوت در شناسایی طغیان‌ها و ۲۵۲ روز بدون رخداد طغیان، با هدف سنجش عملکرد الگوریتم‌های مورد آزمون در تشخیص روزهای بدون طغیان، روی داده‌های پیش‌پردازش شده موارد مشکوک به سرخک، طی سال‌های ۱۳۸۷ تا ۱۳۸۹ ایجاد شد.

روش کشف طغیان

در این مطالعه با استفاده از داده‌های موارد گزارش شده روزانه مشکوک به سرخک، عملکرد الگوریتم مجموع تراکمی که به اختصار $CUSUM$ نام‌گذاری شده است، در شناسایی طغیان‌های

رویکرد ارزشیابی عملکرد الگوریتم کشف طغیان

عملکرد الگوریتم‌های کشف طغیان مجموع تراکمی در شناسایی طغیان‌های نیمه‌شبیه‌سازی شده با استفاده از شاخص‌های حساسیت، ویژگی، درصد هشدار کاذب، درصد منفی کاذب و نسبت‌های درست‌نمایی، به تفکیک نوع طغیان‌های مورد نظر شامل طغیان‌های برگرفته از منابع موجود، طغیان‌های رخ داده سرخک در ایران در سال ۱۳۸۹ و طغیان‌های شبیه‌سازی شده گزارش گردیده است. برای مقایسه عملکرد الگوریتم‌ها از سطح زیر منحنی راک (منحنی خاص محرک گیرنده)، همراه با فاصله اطمینان ۹۵ درصد استفاده شد. برای ارزشیابی به‌هنگام بودن الگوریتم‌ها از شاخص به‌هنگام بودن کشف طغیان ۱، یعنی فاصله بین روز واقعی شروع طغیان (استاندارد طلایی) و نخستین روز یا زمانی پس از آن که الگوریتم اعلام هشدار می‌کند، استفاده شد. همچنین، برای نمایش گرافیکی به‌هنگام بودن الگوریتم‌ها، مقادیر میانگین و انحراف معیار شاخص به‌هنگام بودن کشف طغیان برای انواع طغیان‌های نیمه‌شبیه‌سازی شده گزارش شده است. افزون بر این، نمودار جعبه و میله مربوط به هر الگوریتم، به تفکیک نوع طغیان نیمه‌شبیه‌سازی شده، ترسیم گردید.

یافته‌ها

نتایج عملکرد کلی الگوریتم‌های مجموع تراکمی، یعنی شاخص‌های حساسیت، ویژگی، درصد هشدار کاذب، درصد منفی کاذب و نسبت‌های درست‌نمایی در جدول شماره ۱ نشان داده شده است. همان‌طور که در قسمت روش کار به‌طور کامل شرح داده شد، هریک از الگوریتم‌های مجموع تراکمی، از شماره **CUSUM 1** تا **CUSUM 12** نام‌گذاری شده است که هر یک سطوح آستانه هشدار متفاوتی دارند. هریک از این الگوریتم‌ها روی ۸۳۱ روز طغیان نیمه‌شبیه‌سازی شده و ۲۵۲ روز بدون طغیان آزموده شده‌اند. همان‌طور که در جدول شماره ۱ آمده است، حساس‌ترین روش کشف طغیان در این گروه، الگوریتم **CUSUM 8** با حساسیت ۵۰ درصد و $(CI: 47-54) / 95$ است. بالاترین مقدار ویژگی مربوط به الگوریتم **CUSUM 4** برابر با ۸۷ درصد و $(CI: 83-91) / 95$ است. کمترین مقدار هشدارهای کاذب در الگوریتم **CUSUM 4**، به میزان ۱۳ درصد و $(CI: 9-17) / 95$ و پایین‌ترین درصد منفی کاذب مربوط در الگوریتم **CUSUM 8**، با ۵۰ درصد و $(CI: 46-53) / 95$ نشان داده شده است. بهترین نسبت درست‌نمایی مثبت، یعنی ۲/۲۰، مربوط به الگوریتم **CUSUM 2** و کمترین نسبت درست‌نمایی منفی مربوط به الگوریتم **CUSUM 8** برابر با ۰/۶۸ است.

^۱ Timeliness of detection

(جدول شماره ۱). مقایسه عملکرد کلی الگوریتم‌ها با یکدیگر با استفاده از سطح زیرمنحنی راک بیان‌گر عملکرد بهتر الگوریتم **CUSUM 6**، در مقایسه با دیگر روش‌ها است. به طوری که سطح زیر منحنی مربوط برابر با ۰/۵۹، با $(CI: 0/56-0/63) / 95$ است.

نتایج عملکرد الگوریتم‌های مجموع تراکمی در شناسایی طغیان‌ها، به تفکیک انواع طغیان‌ها، شامل ۴۰۴ روز طغیان برگرفته از منابع موجود، ۱۴۷ روز طغیان نیمه‌شبیه‌سازی شده بر اساس طغیان‌های رخ داده سرخک در ایران، در سال ۱۳۸۹ و ۲۸۰ روز طغیان شبیه‌سازی شده، متفاوت با عملکرد کلی آن در شناسایی کل ۸۳۱ روز طغیان نیمه‌شبیه‌سازی شده است. به طوری که حساس‌ترین روش کشف طغیان در گروه طغیان‌های شبیه‌سازی شده، الگوریتم **CUSUM 8** با حساسیت ۷۸ درصد و $(CI: 73-83) / 95$ است. بالاترین مقدار ویژگی مربوط به الگوریتم **CUSUM 4**، برابر با ۸۷ درصد و $(CI: 83-91) / 95$ است. کمترین مقدار هشدارهای کاذب مربوط به الگوریتم **CUSUM 4**، به میزان ۱۳ درصد و $(CI: 9-17) / 95$ و پایین‌ترین درصد منفی کاذب مربوط به الگوریتم **CUSUM 8**، با ۲۲ درصد و $(CI: 17-27) / 95$ است. برای مقایسه آسان‌تر، منحنی راک مربوط به عملکرد کلی الگوریتم‌ها در شناسایی انواع طغیان‌ها و شناسایی زیرگروه طغیان‌های شبیه‌سازی شده در نمودار شماره ۱ آورده شده است. همان‌طور که دیده می‌شود، بالاترین سطح زیر منحنی راک برای الگوریتم‌های یاد شده در شناسایی انواع طغیان‌ها، به صورت کلی، برابر با ۰/۵۹، با $(CI: 0/56-0/63) / 95$ است، در حالی که این مقدار برای شناسایی زیرگروه طغیان‌های شبیه‌سازی شده برابر با ۰/۷۷، با $(CI: 0/73-0/82) / 95$ است (نمودار شماره ۱).

جدول شماره ۲ میانگین و انحراف معیار شاخص به‌هنگام بودن الگوریتم‌های مجموع تراکمی برای هریک از انواع طغیان‌ها را نشان می‌دهد. به‌طور متوسط، کمترین فاصله زمانی بین روز واقعی شروع طغیان بر مبنای داده‌های نیمه‌شبیه‌سازی شده تا زمان کشف طغیان برای این گروه از الگوریتم‌ها، ۱۳/۸۴ روز است. بیشترین فاصله زمانی بین رخداد تا زمان شناسایی طغیان با این الگوریتم‌ها نیز، به طور متوسط ۱۴/۵۷ روز بود (جدول شماره ۲).

برای مقایسه آسان‌تر میانه، چارک اول، چارک سوم، کمترین و بیشترین مقادیر شاخص به‌هنگام بودن عملکرد الگوریتم‌های مجموع تراکمی، نمودار جعبه و میله در نمودار شماره ۲ ترسیم شده است.

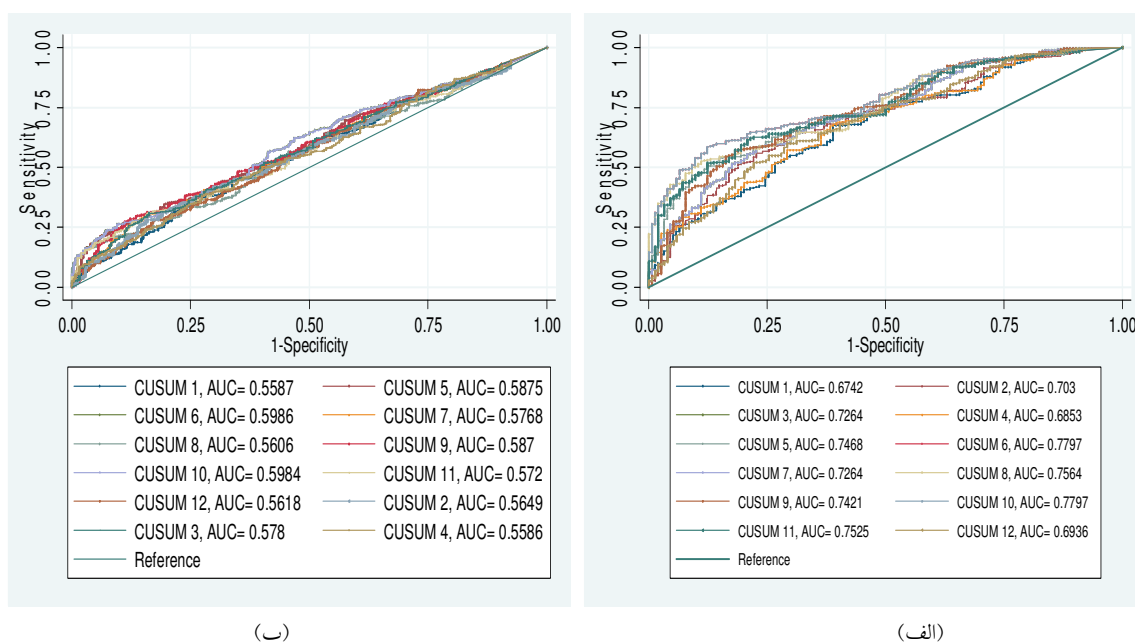
جدول شماره ۱- حساسیت، ویژگی، درصد هشدار کاذب، درصد منفی کاذب و نسبت‌های درست‌نمایی مربوط به الگوریتم‌های مجموع تراکمی در شناسایی انواع طغیان‌های نیمه‌شبیه‌سازی‌شده

الگوریتم	حساسیت (%)	ویژگی (%)	هشدار کاذب (%)	منفی کاذب (%)	نسبت درست‌نمایی مثبت	نسبت درست‌نمایی منفی
CUSUM 1	۳۰ (۳۳-۲۶)*	۸۶ (۹۰-۸۲)	۱۴ (۱۸-۱۰)	۷۰ (۷۴-۶۷)	۲/۱۴	۰/۸۱
CUSUM 2	۳۳ (۳۶-۳۰)	۸۵ (۸۹-۸۱)	۱۵ (۱۹-۱۱)	۶۷ (۷۰-۶۴)	۲/۲۰	۰/۷۹
CUSUM 3	۳۷ (۴۰-۳۴)	۸۲ (۸۷-۷۷)	۱۸ (۲۳-۱۳)	۶۳ (۶۶-۶۰)	۲/۰۶	۰/۷۷
CUSUM 4	۲۷ (۳۰-۲۴)	۸۷ (۹۱-۸۳)	۱۳ (۱۷-۹)	۷۳ (۷۶-۷۰)	۲/۰۸	۰/۸۴
CUSUM 5	۴۱ (۴۴-۳۷)	۸۰ (۸۵-۷۵)	۲۰ (۲۵-۱۵)	۵۹ (۶۳-۵۶)	۲/۰۵	۰/۷۴
CUSUM 6	۴۶ (۵۰-۴۳)	۷۵ (۸۱-۷۰)	۲۵ (۳۰-۱۹)	۵۴ (۵۷-۵۰)	۱/۸۴	۰/۷۲
CUSUM 7	۳۷ (۴۰-۳۴)	۸۳ (۸۷-۷۸)	۱۷ (۲۲-۱۳)	۶۳ (۶۶-۶۰)	۲/۱۸	۰/۷۶
CUSUM 8	۵۰ (۵۴-۴۷)	۷۳ (۷۸-۶۷)	۲۷ (۳۳-۲۲)	۵۰ (۵۳-۴۶)	۱/۸۵	۰/۶۸
CUSUM 9	۳۹ (۴۲-۳۶)	۸۲ (۸۷-۷۷)	۱۸ (۲۳-۱۳)	۶۱ (۶۴-۵۸)	۲/۱۷	۰/۷۴
CUSUM 10	۴۶ (۵۰-۴۳)	۷۵ (۸۱-۷۰)	۲۵ (۳۰-۱۹)	۵۴ (۵۷-۵۰)	۱/۸۴	۰/۷۲
CUSUM 11	۴۳ (۴۶-۴۰)	۷۴ (۷۹-۶۸)	۲۶ (۳۲-۲۱)	۵۷ (۶۰-۵۴)	۱/۶۵	۰/۷۷
CUSUM 12	۳۵ (۳۸-۳۱)	۸۲ (۸۷-۷۷)	۱۸ (۲۳-۱۳)	۶۵ (۶۹-۶۲)	۱/۹۴	۰/۷۹

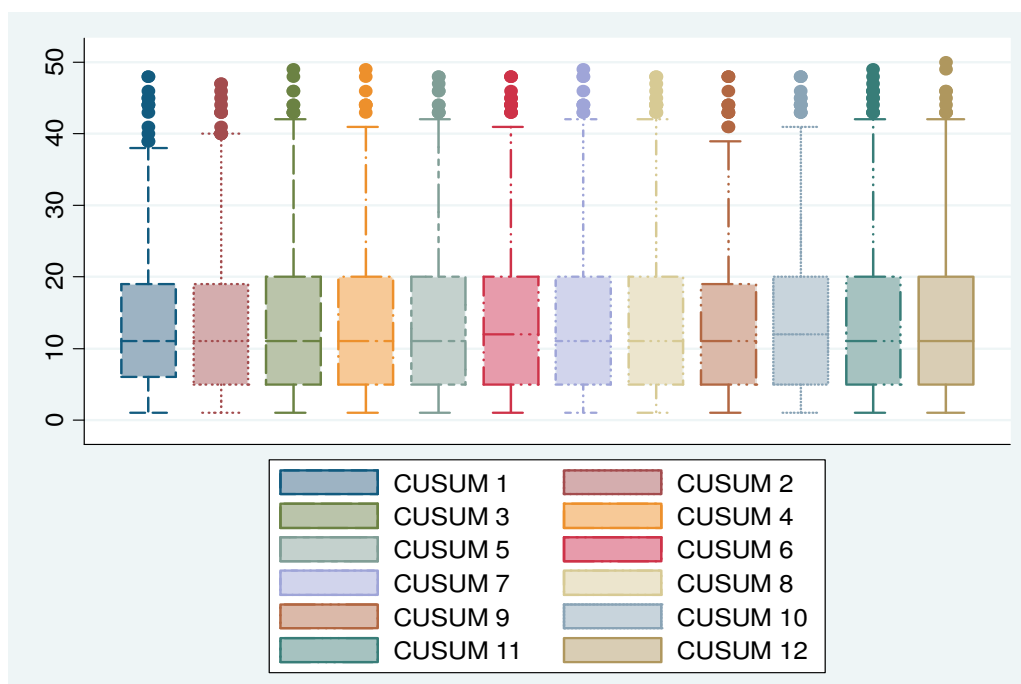
* اعداد داخل پرانتز فاصله اطمینان ۹۵ درصد را نشان می‌دهند.

جدول شماره ۲ - میانگین و انحراف معیار شاخص به‌هنگام‌بودن الگوریتم مجموع تراکمی در شناسایی طغیان‌های نیمه‌شبیه‌سازی شده، به تفکیک نوع طغیان

الگوریتم	کل طغیان‌ها		طغیان‌های برگرفته از منابع موجود		طغیان‌های سرخک در ایران		طغیان‌های شبیه‌سازی شده	
	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار
CUSUM 1	۱۴/۱۳	۱۱/۴۰	۲۱/۸۵	۱۳/۰۵	۱۸/۷۸	۱۳/۵۴	۸/۶۷	۵/۴۳
CUSUM 2	۱۳/۸۴	۱۱/۲۱	۲۱/۹۹	۱۲/۶۳	۱۷/۰۰	۱۳/۳۱	۸/۵۸	۵/۶۵
CUSUM 3	۱۴/۱۱	۱۱/۳۳	۲۲/۲۲	۱۲/۸۱	۱۸/۹۱	۱۲/۷۷	۸/۶۲	۵/۵۸
CUSUM 4	۱۴/۱۹	۱۱/۴۲	۲۲/۲۲	۱۲/۹۸	۱۷/۸۱	۱۱/۸۹	۸/۵۲	۵/۵۱
CUSUM 5	۱۴/۱۳	۱۱/۴۲	۲۲/۲۶	۱۲/۷۸	۱۸/۹۶	۱۳/۱۰	۸/۵۵	۵/۶۳
CUSUM 6	۱۴/۵۷	۱۱/۶۵	۲۲/۲۵	۱۲/۷۶	۱۸/۹۱	۱۲/۹۲	۸/۵۰	۵/۶۳
CUSUM 7	۱۴/۱۶	۱۱/۳۹	۲۲/۲۹	۱۲/۸۹	۱۸/۹۱	۱۲/۷۷	۸/۶۲	۵/۵۸
CUSUM 8	۱۴/۴۳	۱۱/۴۵	۲۱/۵۸	۱۲/۶۰	۱۸/۴۷	۱۲/۵۰	۸/۴۳	۵/۵۵
CUSUM 9	۱۴/۰۲	۱۱/۳۲	۲۲/۱۵	۱۲/۷۸	۱۷/۵۲	۱۳/۳۶	۸/۶۴	۵/۵۷
CUSUM 10	۱۴/۵۷	۱۱/۶۵	۲۲/۲۵	۱۲/۷۶	۱۸/۹۱	۱۲/۹۲	۸/۵۰	۵/۶۳
CUSUM 11	۱۴/۴۴	۱۱/۶۲	۲۲/۴۸	۱۲/۸۸	۱۹/۲۴	۱۳/۲۴	۸/۶۰	۵/۶۴
CUSUM 12	۱۴/۱۳	۱۱/۳۵	۲۲/۳۴	۱۲/۹۰	۱۸/۸۰	۱۲/۸۹	۸/۶۹	۵/۶۰



نمودار شماره ۱- منحنی راک مرتبط با الگوریتم‌های مجموع تراکمی در شناسایی انواع طغیان‌ها (الف) و زیرگروه طغیان‌های شبیه‌سازی شده (ب)



نمودار شماره ۲ - نمودار جعبه‌ای شاخص به هنگام بودن الگوریتم مجموع تراکمی در شناسایی انواع طغیان‌های نیمه‌شبیه‌سازی شده

بحث

بهتر عملکرد الگوریتم مجموع تراکمی در شناسایی این گروه از طغیان‌ها منجر شده است. یافته اخیر با نتایج مطالعه هوت‌واگنر و همکارانش (۳۱) سازگار است.

نتایج مقایسه عملکرد کلی الگوریتم‌ها با استفاده از سطح زیر منحنی راک بیانگر کارایی بهتر الگوریتم مجموع تراکمی در شناسایی طغیان‌های شبیه‌سازی شده است. این یافته‌ها با نتایج مطالعه باکریج و همکارانش (۱۱) همخوان است و با مطالعه جکسون و همکارانش (۳۰) که عملکرد بهتر مدل خطی تعمیم یافته را بیان می‌کند، ناهمخوان است. یکی از مهم‌ترین دلایل ناهمخوانی نتایج، پیش پردازش نشدن داده‌ها و حذف الگوهای توجیه‌پذیر در مطالعه جکسون و همکارانش است. به طوری که نقش آثار روزهای هفته فقط در مدل خطی تعمیم یافته در نظر گرفته شده بود.

مقایسه عملکرد الگوریتم‌ها از نظر به‌هنگام‌بودن، با استفاده از شاخص به‌هنگام‌بودن کشف طغیان‌ها، با نتایج مطالعه جکسون و همکارانش (۳۰) به دلیل در نظر نگرفتن آثار روزهای هفته در فرآیند ارزشیابی الگوریتم مجموع تراکمی، همخوانی ندارد. با این وجود، از نظر شاخص حساسیت شناسایی کلی طغیان‌ها، بدون توجه به زمان کشف و به‌هنگام‌بودن آن‌ها، عملکرد الگوریتم مجموع تراکمی بهتر است.

داوری درباره کارایی روش‌های کشف طغیان و نظام‌های مراقبت سندرمیک فقط تابع شاخص‌های حساسیت، ویژگی، درصد هشدارهای کاذب و به‌هنگام بودن روش کشف طغیان مورد استفاده نیست. بلکه به کیفیت داده‌ها و پاسخ به‌هنگام به طغیان شناسایی شده نیز، وابسته است. یافته‌های مطالعه حاضر درباره عملکرد بهتر الگوریتم‌ها در شناسایی طغیان‌های با اندازه بزرگ‌تر با نتایج مطالعه‌های مشابه (۲۷-۲۹) همخوان است. افزون بر این، عملکرد مناسب الگوریتم‌ها در شناسایی طغیان‌هایی که به روش شبیه‌سازی شده به داده‌ها اضافه شده بودند، بیانگر کارایی آن‌ها در کشف آن‌دسته از طغیان‌هایی است که شکل منحنی همه‌گیری آن‌ها موج‌های متوالی و بدون فاصله دارد. وجود منحنی‌های همه‌گیری چندمنمایی در طغیان‌های برگرفته از نظام مراقبت سرخک در سال ۱۳۸۹، در ایران تأییدکننده این یافته در مطالعه حاضر است. از این رو، شکل منحنی همه‌گیری طغیان‌های افزوده به داده‌های مطالعه، نقشی مهم در شاخص به‌هنگام‌بودن کشف طغیان‌ها دارد. یافته اخیر با نتایج مطالعه جکسون و همکارانش (۳۰) و نیز مطالعه ونگ و همکارانش (۲۷) همخوانی دارد. از سویی، کوچک بودن اندازه طغیان‌ها و فاصله‌های طولانی بین منحنی همه‌گیری در طغیان‌های برگرفته از منابع موجود و طغیان‌های برگرفته از نظام مراقبت سرخک در ایران، به کارایی

ویژگی‌های خاص کشور ایران را دارد. در نتیجه، مقایسه نتایج مطالعه حاضر را دچار مشکل می‌کند. با این وجود، در این مطالعه تلاش شده است تا اطلاعاتی جامع درباره داده‌های مورد استفاده، شامل طغیان‌های افزوده به آن‌ها و فرآیند پیش‌پردازش داده‌ها، به تفصیل بیان شود تا امکان مقایسه نتایج با دیگر مطالعه‌ها فراهم گردد.

نتیجه‌گیری

به‌طور کلی می‌توان چنین نتیجه‌گیری کرد که عملکرد الگوریتم کشف طغیان مورد استفاده در شناسایی طغیان‌های نیمه‌شبیه‌سازی شده، ضعیف است. با این وجود، داوری درباره کارایی روش‌های مورد استفاده در این مطالعه باید با در نظر گرفتن نوع طغیان‌های به‌کاررفته، از نظر شکل، اندازه و دوره طغیان، در مقایسه با دیگر مطالعه‌های همانند، صورت گیرد.

تشکر و قدردانی

نویسندگان از یاری‌های دست‌اندرکاران نظام مراقبت از بیماری سرخک در کشور و همکاری جناب آقای دکتر زهرایی و سرکار خانم صبوری در استخراج داده‌های موارد گزارش شده مشکوک به سرخک، صمیمانه سپاسگزارند.

از آشکارترین نقاط قوت مطالعه حاضر می‌توان به استفاده همزمان از حالت‌های مختلف یک روش کشف طغیان، با در نظر گرفتن دوره‌های زمانی متفاوت برای ایجاد سطوح آستانه متنوع و در نتیجه، دستیابی به مناسب‌ترین الگوریتم اشاره کرد. افزون بر این، می‌توان به نمایش گرافیکی عملکرد الگوریتم‌ها با استفاده از سطح زیر منحنی راک و نیز نمایش گرافیکی شاخص به‌هنگام‌بودن کشف طغیان‌ها، با استفاده از نمودار جعبه و میله اشاره کرد. شناسایی و حذف روندهای فصلی و الگوهای توجیه‌پذیر از داده‌های پیش‌از آزمون الگوریتم‌ها و استفاده از طغیان‌های واقعی، به‌جای اکتفا به آزمون الگوریتم‌ها روی طغیان‌های صرفاً شبیه‌سازی شده، از دیگر نقاط قوت مطالعه حاضر است که مطالعه‌های اندکی (۲۶) در دنیا، به‌طور همزمان چنین متدولوژی‌ای را در ارزشیابی روش‌های کشف طغیان به‌کار گرفته‌اند.

درباره محدودیت‌های مطالعه حاضر نیز می‌توان به این نکته اشاره کرد که اگرچه در این مطالعه از طیفی گسترده از طغیان‌ها استفاده شده و تلاش شده است که نمایانگر طغیان‌های واقعی و ممکن‌الوقوع در کشور ایران باشند، ولی نمی‌توان تضمین کرد که همه طغیان‌های احتمالی بیماری سرخک را پوشش داده است. همچنین، همانند دیگر مطالعه‌های مشابه با رویکرد نیمه‌شبیه‌سازی، داده‌های این مطالعه که طغیان‌های شبیه‌سازی شده به آن اضافه شده‌اند، متفاوت با دیگر مطالعه‌ها است و

منابع

- Chen H, Zeng D, Yan P. Infectious Disease Informatics Syndromic Surveillance for Public Health and Bio Defense. 1st ed. New York: Springer Science and Business Media; 2010.
- Zeng D, Chen H, Chavez C, Iober WB, Thurmond MC. Infectious Disease Informatics and Biosurveillance. 1st ed. New York: Springer Science and Business Media; 2011.
- CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. . MMWR 2004; 53(RR-5).
- Henning KJ. What is Syndromic Surveillance? In: Syndromic Surveillance: Reports from a National Conference, 2003. MMWR. 2004; 53: 7-11.
- Wagner MM, Moore AW, Aryel RM, eds. Handbook of Biosurveillance. 1st ed. Oxford: Elsevier Academic Press; 2006.
- Lombardo J, Burhom H, Elbert E, Magruder S, Lewis S, Loschen W, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health. 2003; 80: 32-42.
- Bradley CA, Rolka H, Walker D, Loonsk J. BioSense: implementation of a National Early Event Detection and Situational Awareness System. MMWR Morb Mortal Wkly Rep. 2005; 54: 11-9.
- Aamodt G, Samuelsen SO, Skrondal A. A simulation study of three methods for detecting disease clusters. Int J Health Geogr. 2006; 5: 15.
- Siegrist D, Pavlin J. Bio-ALIRT biosurveillance detection algorithm evaluation. MMWR Morb Mortal Wkly Rep. 2004; 53: 152-8.
- Kleinman KP, Abrams A, Mandl K, Platt R. Simulation for assessing statistical methods of biologic terrorism surveillance. MMWR Morb Mortal Wkly Rep. 2005; 54: 101-8.
- Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. J Biomed Inform. 2005; 38: 99-113.
- World Health Organisation. Measles. 2011; Available from: <http://www.who.int/immunization/topics/measles/en/index.html>.
- Wolfson LJ, Strelbel PM, Gacic-Dobo M, Hoekstra EJ, McFarland JW, Hersh BS. Has the 2005 measles mortality reduction goal been achieved? A natural history modelling study. Lancet. 2007; 369: 191-200.
- Sahni V, Naus M, Preston E, Kidd C, Miller L, Guy C, et al. Health Care Worker (HCW) Exposures in Hospitals During a Measles Outbreak. Canadian Public Health Association Conference; June 18 to 22nd, 2011.
- Sugerman DE, Barskey AE, Delea MG, Ortega-Sanchez IR, Bi D, Ralston KJ, et al. Measles outbreak in a highly vaccinated population, San Diego, 2008: role of the intentionally undervaccinated. Pediatrics. 2010; 125: 747-55.
- Marinova L, Kojouharova M, Mihneva Z. An ongoing measles outbreak in Bulgaria, 2009. Euro Surveill. 2009; 14.
- Gupta SN, Gupta N. Two Highly Immunized Hilly Areas versus Double Measles Outbreak Investigations in District Kangra, Himachal Pradesh, India, in 2006. J Glob Infect Dis. 2009; 1: 14-20.

18. Seguliev Z, Duric P, Petrovic V, Stefanovic S, Cosic G, Hrnjakovic IC, et al. Current measles outbreak in Serbia: a preliminary report. *Euro Surveill.* 2007; 12: E070315 2.
19. Filia A, Barale A, Malaspina S, Montu D, Zito S, Muscat M, et al. A cluster of measles cases in northern Italy: a preliminary report. *Euro Surveill.* 2007; 12: 3318.
20. Yeung LF, Lurie P, Dayan G, Eduardo E, Britz PH, Redd SB, et al. A limited measles outbreak in a highly vaccinated US boarding school. *Pediatrics.* 2005; 116: 1287-91.
21. Rooney JA, Milton DJ, Hackler RL, Harris JH, Reynolds D, Tanner M, et al. The largest outbreak of measles in the United States during 1999: imported measles and pockets of susceptibility. *J Infect Dis.* 2004; 189: 78-80.
22. Lofgren JP, Cochran V, Abbott O, Woernle C. Measles Outbreak Associated with an Imported Case in an Infant, Alabama, 2002. *MMWR.* 2004; 53: 30-3.
23. van den Hof S, Meffre CM, Conyn-van Spaendonck MA, Woonink F, de Melker HE, van Binnendijk RS. Measles outbreak in a community with very low vaccine coverage, the Netherlands. *Emerg Infect Dis.* 2001; 7: 593-7.
24. Page ES. Continuous Inspection Schemes. *Biometrika.* 1954; 41: 100-15.
25. Karami M, Soori H, Mehrabi Y, Haghdoost AA, Gouya MM. Detecting and removing the explainable patterns of the daily counts of suspected cases of measles as a prediagnostic data source in Iran. *Iranian Journal of Epidemiology;* 2012.
26. Karami M. Evaluating the performance of outbreak detection algorithms in surveillance systems using semi-synthetic approach [PhD thesis]. Tehran: Shahid Beheshti University of Medical Sciences; 2012.
27. Wang L, Ramoni MF, Mandl KD, Sebastiani P. Factors affecting automated syndromic surveillance. *Artif Intell Med.* 2005; 34: 269-78.
28. Hutwagner L, Browne T, Seeman GM, Fleischauer AT. Comparing aberration detection methods with simulated data. *Emerg Infect Dis.* 2005; 11: 314-6.
29. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak.* 2003; 3: 2.
30. Jackson ML, Baer A, Painter I, Duchin J. A simulation study comparing aberration detection algorithms for syndromic surveillance. *BMC Med Inform Decis Mak.* 2007; 7: 6.
31. Hutwagner LC, Thompson WW, Seeman GM, Treadwell T. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. *Stat Med.* 2005; 24: 543-50.

Evaluating the Performance of an Outbreak-Detection Algorithms using Semi-Synthetic Approach: Cumulative Sum Algorithm

Karami M¹, Soori H², Mehrabi Y³, Haghdoost AA⁴, Gouya MM⁵, Esmailnasab N⁶

1- Assistant Professor of Epidemiology, Research Center for Modeling of Noncommunicable Diseases and Department of Biostatistics & Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

2- Professor of Epidemiology, Safety Promotion and Injury Prevention Research Center, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3- Professor of Biostatistics, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

4- Professor of Epidemiology and Biostatistics, Research Centre of Modeling in Health, Institute of futures studies in health, Kerman University of Medical Sciences, Kerman, Iran

5- Director General, Center for Disease Control, Ministry of Health & Medical Education, Tehran, Iran

6- Associate Professor of Epidemiology, Department of Epidemiology and Biostatistics, Kurdistan University of Medical Sciences, Kurdistan, Iran

Corresponding author: Karami M., ma.karami@umsha.ac.ir

Background & Objectives: Timely response to emerging diseases and outbreaks are a major public health and health systems priority. There are few published studies that evaluate the performance of cumulative sum (CUSUM) on identical data using semi-synthetic simulation approach. This study was undertaken to determine the performance of the CUSUM in timely detection of 831 days of simulated outbreaks.

Methods: We evaluated the performances of the CUSUM as an outbreak detection method on simulated outbreaks injected to daily counts of suspected cases of measles as baseline data in Iran between 21 March 2008 till 20 March 2011. Data obtained from the Iranian national surveillance system. The performance of algorithms was evaluated using sensitivity, false alarm rate, likelihood ratios and Area under the Receiver Operating Characteristic (ROC) curve.

Results: Generally the sensitivity of the CUSUM algorithm in detecting simulated outbreaks was 50% (95% CI: 47- 54). The corresponding values are disaggregated according to outbreak size, shape and duration. The CUSUM algorithm detected the half of outbreaks after 13.84 days on average.

Conclusion: We concluded that CUSUM algorithm performed good in detection of large outbreaks with short periods and poorly in detecting long period outbreaks, particularly those simulated outbreaks that did not begin with a surge of cases.

Keywords: Surveillance System, Measles, Outbreak Detection Method, CUSUM, Semi-synthetic Simulation, Iran