

TITLE:

**Influenza and other respiratory viruses:** standardizing disease severity in surveillance and clinical trials

ABSTRACT:

**Introduction:** Influenza-Like Illness is a leading cause of hospitalization in children. Disease burden due to influenza and other respiratory viral infections is reported on a population level, but clinical scores measuring individual changes in disease severity are urgently needed. **Areas covered:** We present a composite clinical score allowing individual patient data analyses of disease severity based on systematic literature review and WHO-criteria for uncomplicated and complicated disease. **The 22-item ViVI Disease Severity Score showed a normal distribution in a pediatric cohort of 6073 children aged 0–18 years (mean age 3.13; S.D. 3.89; range: 0 to 18.79).** **Expert commentary:** The ViVI Score was correlated with risk of antibiotic use as well as need for hospitalization and intensive care. The ViVI Score was used to track children with influenza, respiratory syncytial virus, human metapneumovirus, human rhinovirus, and adenovirus infections and is fully compliant with regulatory data standards. The ViVI Disease Severity Score mobile application allows physicians to measure disease severity at the point-of care thereby taking clinical trials to the next level.

**Introduction:**

Influenza-like illness (ILI) and acute respiratory infections (ARI) in children are common. The clinical presentation may range from subtle to severe symptoms requiring advanced medical care [1,2]. The wide spectrum of disease presentations and the role of risk factors (RFs) in terms of disease severity are poorly understood. Laboratory diagnostics are not usually ordered in routine care [3–6].

Surveillance programs should rely on laboratory-confirmed cases rather than clinical suspicion to solve the denominator problem. This will allow the timely detection of virus-specific seasonality in a given (sub)population [7].

An even greater challenge will present itself when investigators wish to determine the impact of different respiratory viruses on disease burden [8]. A deeper understanding of disease severity in relation to specific respiratory viruses will help in the monitoring of the real-world impact of ‘natural’ or untreated disease as well as preventive measures and therapeutic interventions such as vaccines and antivirals. The timely detection of seasonality will help with the targeted and cost-effective use of viral diagnostics in hospital-based surveillance settings. Ideally, viral diagnostics should be aligned with simultaneous standardized disease severity assessments.

Standardized measures of disease severity are urgently needed for clinical trials of vaccines and antivirals currently in development for ARI caused by influenza (FLU), respiratory syncytial virus (RSV), human metapneumovirus (HMPV), adenovirus (ADV), or human rhinovirus (HRV) [9–17]. Furthermore, it would be desirable to assess, at the point of care, which patients are suffering from severe disease in relation to their perceived RF profile, and to use such point-of-care assessments to individualize the use of anti-infective therapy. Experience during the recent influenza pandemic has shown that influenza disease severity appears rather unpredictable, especially in young patients. Whilst the majority of adults with severe disease did have previously identifiable RFs, the majority of children affected by severe disease did not [18,19]. The expected or perceived risk of severe outcomes may also influence a physician’s decision to test a patient for influenza and other respiratory viruses [20]. There is little consensus on which symptoms should trigger a physician’s suspicion, and local practices differ significantly from site to site and from season to season [19,21–24].

Comprehensive reviews of the published literature and disease severity measurements used in clinical trials and surveillance systems are lacking. The numerous observational studies and clinical trials assessing the prevention and treatment influenza and other respiratory viruses have been rather inconsistent. Commonly used indicators of disease severity such as ‘hospitalization,’ a diagnosis of ‘pneumonia,’ and other adverse outcomes including mortality are known to be highly dependent on the studied population, the medical setting, the choice of data sources, and the availability of resources [25]. Head-to-head comparisons and meta-analyses comparing different preventive and treatment interventions will require universally accepted disease severity measurements.

Sentinel surveillance systems tend to focus on private practices and laboratory testing based on clinical suspicion on behalf of primary care providers working at surveillance sites [26]. With children being the most prominent transmitters of influenza, pediatric emergency rooms and large tertiary care hospitals are ideal sites to monitor seasonality covering the entire spectrum of clinical presentations [27–29]. To create a model system free of selection bias, a perennial quality management (QM) program was instituted at a large pediatric academic center in collaboration with the National Reference Centre for Influenza and Other Respiratory Viruses [30–34].

The specific aims of the presented analyses are

to develop a standardized approach to measuring ILI disease severity based on literature review and WHO guidelines and to apply new mathematical models to the real-time surveillance of ILI in large tertiary care centers.

#### Literature review ::: Methods:

To understand which disease severity parameters have been used in clinical trials and observational studies, a systematic literature search of the PubMed database was performed using the following search terms: '(disease severity[Title/Abstract] OR illness severity[Title/Abstract]) AND (influenza[Title/Abstract] OR rhinoviruses[Title/Abstract] OR adenovirus[Title/Abstract] OR human metapneumovirus[Title/Abstract] OR Respiratory Syncytial Virus[Title/Abstract] OR Coronavirus[Title/Abstract] OR bocavirus[Title/Abstract] OR parainfluenza virus[Title/Abstract] OR respiratory virus[Title/Abstract]).' For the purposes of this expert review, the literature review was updated covering publications dating from 1 January 2006 to 8 June 2016. Searches were limited to human studies published in English. Abstracts were screened manually and excluded according to the following criteria: (1) studies were not pediatric or study subjects were, in the majority, >18 years old; (2) studies were not one of the following: randomized clinical trials, non-randomized clinical trials, observational studies, or epidemiological studies; and (3) studies lacked any clinical criterion for disease severity. Animal studies, adult studies, meta-analysis, and review papers were also excluded.

#### The ViVI Disease Severity Score ::: Methods:

Based on the systematic literature review, the ViVI Disease Severity Score was developed as a 22-item weighed clinical composite score, according to WHO-criteria of uncomplicated and complicated disease [35]. The ViVI Disease Severity Score is comprised of 9 items describing signs and symptoms of uncomplicated disease (Disease Severity, Uncomplicated: DSU, weighed single-fold) reflecting 'regular' ILI activity, whereas the 13 items describing parameters consistent with complicated disease (Disease Severity, Complicated: DSC, weighed threefold) indicate high-impact clinical presentations in the target population (Textbox 1). The ViVI Disease Severity Score was subsequently user tested as a web-user interface as well as a mobile application for tablet computers, to be used at the point of care.

#### The ViVI Risk Factor Score ::: Methods:

Based on the 16 most commonly cited RFs for severe disease in the pediatric or adolescent age group, a simple RF score was composed [35–38]. The ViVI Risk Factor Score (Textbox 2) was implemented on the same mobile application to allow the reporting of disease severity in relation to previously identifiable RFs in the individual patient.

#### The Consultation Index ::: Methods:

The Consultation Index is an epidemiological indicator reported weekly by the National Reference Centre for Influenza and Other Respiratory Viruses and the Influenza Working Group, based on the proportion of ARI at representative sentinel practices across the country [39].

The Consultation Index represents a timely indicator of any deviation from a baseline rate of ARI cases presenting to the respective sentinel practices. A 'normal ARI activity' is assumed if the Consultation Index remains below 115. Increased activities are typically measured during the winter months, when seasonal viruses circulate in the community.

Fluctuations in ARI activity as measured by the Consultation Index represent a useful indicator of disease burden based on actual case numbers. Reporting of the number of cases, however, does not reveal information on disease severity with each individual case. By plotting the Consultation Index with the corresponding average ViVI Disease Severity Score in the same graph, we obtain a comprehensive picture of ARI disease burden that is based on both actual case numbers and

case severity. Figure 4(a) illustrates that disease severity does not always follow the peaks and troughs of case numbers as measured using the Consultation Index [39]. The ViVI Score and the Consultation Index are therefore measuring opposing end points; one is based on individual disease severity per patient (ViVI Disease Severity Score) and the other serves as an epidemiological indicator of ARI activity and the overall disease burden within the national surveillance system (Consultation Index).

#### Cohort design and patient population :: Methods:

The ViVI Disease Severity and Risk Factor Score were user tested in the context of a QM program for children with ILI at a large pediatric hospital in Germany as described previously [30–34]. According to the standard operating procedures, patients with a physician diagnosis of ILI and/or fulfilling predefined case criteria (body temperature  $\geq 38^{\circ}\text{C}$  and  $\geq 1$  respiratory symptom) admitted to the emergency department (ED) or pediatric inpatient wards, participated in the QM program [30–34]. Independent of routine clinical care, a specifically trained QM team obtained nasopharyngeal samples and performed standardized clinical assessments using the ViVI Disease Severity Score in line with WHO criteria for uncomplicated and complicated influenza [30,35,40]. The ViVI Disease Severity Score was recorded at the first consultation with patients participating in the QM Program. Physicians in routine care were blinded to the results of the clinical assessments by QM staff, and they were unaware of the ViVI Disease Severity Scores assigned by the QM team. QM staff on the other hand assessed patients prior to allocation and treatment decisions on behalf of the clinical team in routine care [30]. Nasopharyngeal specimens were delivered to the National Reference Centre for Influenza and Other Respiratory Viruses for individual RT-PCR testing influenza virus A and B, RSV, HMPV, HRV, and ADV as described below. From December 2009 until April 2015, a total of 6073 children aged 0–18 years participated in the QM program. The QM program included both in- and outpatients to represent the broadest possible spectrum of disease severity. From 2009 to 2015, all patients presenting the ED were screened for ILI criteria once weekly, regardless of whether they were subsequently admitted to the hospital or not. From 2011 onward, daily screenings of all inpatients were added (including weekends and holidays). The QM team performed the disease severity assessments independently and the results remained unknown to the routine staff. Hence, the data acquired by the QM team did not have any influence on treatment or hospitalization decisions. Also, the treating physician did not know the result of the RT-PCR testing when deciding on neuraminidase inhibitor treatment.

Patients with laboratory-confirmed influenza infection were invited to participate in follow-up assessments whenever feasible. Follow-up visits in the QM program were voluntary and scheduled according to the parent's preferences. During follow-up visits, the ViVI Disease Severity Score assessment was repeated and recorded by the QM team using the same procedure as during the initial assessment. Nasopharyngeal samples were repeated and sent for analogous RT-PCR testing [32]. The QM program was approved by the Institutional Review Board (EA 24/008/10). Informed consent procedures were waived for enhanced quality of care and infection control [30–34,40].

#### Laboratory methods :: Methods:

Nasopharyngeal swabs were washed out in a total volume of 3 ml of cell culture medium either individually or pooled per patient. RNA was extracted from 300  $\mu\text{l}$  of patient specimen using the MagAttract Viral RNA M48 Kit (Qiagen, Hilden, Germany) and eluted in 80  $\mu\text{l}$  elution buffer. Alternatively, RNA was extracted using the MagNA Pure 96 DNA and Viral NA Small Volume Kit (Roche Deutschland Holding GmbH, Mannheim, Germany) from 200  $\mu\text{l}$  specimen with an elution volume of 50  $\mu\text{l}$ . A volume of 25  $\mu\text{l}$  of extracted RNA was subjected to cDNA synthesis applying 200 U M-MLV Reverse Transcriptase (Invitrogen, Karlsruhe, Germany) in a total volume of 40  $\mu\text{l}$ . All cDNA samples were analyzed by RT-PCR for the presence of each of the pathogens influenza virus A and B, RSV, HMPV, HRV, and ADV as published previously [41–45].

#### Statistical analysis :: Methods:

A descriptive analysis of the study sample was performed by calculating proportions and summarizing continuous variables using mean (standard deviation and range) and median (interquartile range). Histograms and box plots were used to illustrate the distribution of ViVI Disease Severity Scores. Correlations between the ViVI Disease Severity Score and the Consultation Index were assessed using scatter plots and Pearson's correlation coefficient. The mean difference in ViVI Disease Severity Scores was compared across patient and clinical

characteristics. Statistical significance was assessed using the t-test or the chi-squared test as appropriate. To test whether patients with elevated ViVI Disease Severity Scores also had elevated RF scores, we performed correlation analysis using Pearson's correlation coefficient. These analyses were conducted using Stata version 14 (Statacorp LP, Texas, USA).

We further performed regression analyses to identify a set of influential RFs that could model a linear correlation: ViVI Disease Severity Score =  $w_1 \times \text{RF } 1, w_2 \times \text{RF } 2, \dots, w_n \times \text{RF } n$ . Here,  $w_i$  is the respective weight factor for feature  $i$  in the regression model [46].

In a subset of patients with laboratory-confirmed influenza infection during the 2011/12 and 2012/13 winter seasons, the ViVI Disease Severity Score was also used to follow patients longitudinally with respect to viral load and disease severity over time [32]. To assess the relationship between ViVI Disease Severity Score and virus load, we performed Pearson correlation analyses for all records, for which more than two follow-up time point with virology and ViVI Disease Severity Scores was available. Decision tree analysis [47] was performed to study the relationship between subgroups with a strong positive and negative correlation between disease severity and virus load.

Time series analysis with change point detection ::: Methods:

As an objective and data-driven measure to detect seasonality of respiratory viral infections in acute care settings, we introduced time series analysis with change point (CP) detection. The goal of CP detection algorithms is to identify changes in the dynamical behavior within a time series [48]. The main difference to a statistically oriented analysis is that it assumes that an intrinsic dynamics model generates the data. CP detection therefore identifies those time points, when time series trends start differing significantly from previous data. This procedure allows identification of critical time points when weekly average numbers of laboratory-confirmed influenza infections start to increase (or decrease) compared to preceding weeks. For further detail on CP detection, please refer to the Supplemental Data.

In this paper, we used the CP detection approach to analyze the QM dataset, which allowed computing averages of target variables assigned to respective calendar weeks (such as average rates of laboratory-confirmed influenza infections per calendar week, average disease severity per calendar week, etc.).

For the detection of seasonal patterns, we used the following three-step algorithm. (1) The data were clustered using k-means clustering [49] into potential seasons. We used  $k = 3$  to model two main seasons (high and low) and a transition between those seasons. (2) We assigned a preliminary CP to a week  $wt$ , if the cluster assignment  $c(wt)$  to the respective week differed from the cluster assignment to the preceding week, i.e. if  $c(wt) \neq c(wt-1)$ . (3) Finally, we computed a list of preliminary CPs that would split the dataset into time frames  $tf_1 \dots tf_n$ , where each time frame  $tf_i$  was defined to lie between two consecutive CPs. We then checked for each preliminary CP, whether the values before and after the CP (for the two time frames  $tf_{i-1} \dots tf_i$ ) differed significantly ( $p < 0.05$ ) based on a t-test. This procedure ensures that two regions separated by a CP are indeed different. All preliminary CPs fulfilling the above criteria were reported.

Literature review ::: Results:

The systematic literature search yielded 613 potentially relevant articles. Among these, 529 articles were excluded based on the criteria mentioned above. An additional 56 studies lacked specific criteria for disease severity. Finally, a total number of 84 eligible articles were identified, the characteristics of which are summarized in Textbox 3.

It became evident that several clinical parameters were shared by multiple studies, as for example hospitalization, oxygen requirement, labored breathing, (P)ICU admission, mortality, feeding problems/dehydration/vomiting, fever, wheezing or abnormal breath sounds, etc. All of these commonly used criteria were included in the ViVI Disease Severity Score (see also: Tief et al. [30], in Textbox 3) except for mortality, which is usually recorded separately in hospital records.

Patient baseline demographics and hospital course ::: Results:

The ViVI Disease Severity Score was validated in the full QM Cohort comprised of 6073 patients aged 0–18 years (mean: 3.13 years; SD: 3.89; range: 0–18.79 years). A percentage of 33.6 of the QM program participants was under the age of 1 year, 51.0% were aged 1–5 years, 13.6% were in the age group 6–15 years, and 1.8% were aged 16–18 years. A total of 3399 (56.0%) of the participants were male. A total of 1685 (27.8%) participants were prescribed antibiotics while only 202 (3.3%) were prescribed antivirals in hospital.

At presentation, 3172 (52.2%) were assessed as being in need of hospitalization with 997 (16.4%) being in need of intensive care (including assisted ventilation and extracorporeal membrane oxygenation). With regard to viral etiology, in decreasing order of frequency, we identified rhinovirus (22.9%), RSV (17.2%), ADV (9.6%), A(H1N1) influenza virus (4.5%), metapneumovirus (4.4%), A(H3N2) influenza virus (2.8%), influenza B viruses of the Victoria-lineage influenza (1.7%), and type B viruses of the Yamagata-lineage (2.0%). In 5.8% of the cases, there was more than 1 concurrent viral infection. Table 1 summarizes the findings from the RF assessment exercise carried out as part of the quality monitoring and Table 2 summarizes the clinical symptoms at presentation.

A total of 702 patients (11.6%) had chest-radiography in the ED. Chest radiography findings showed that 438 (7.2%) had pneumonia, 84 (1.4%) had bronchitis, 1 (0.02%) had bronchiectasis, 3 (0.05%) had bronchiolitis, and 33(0.5%) had other non-pneumonia abnormalities. One hundred and nineteen (2.0%) had a lumbar puncture done in the ED and 113 (1.9%) had cerebrospinal fluid (CSF) chemistry done. Sixty-nine (1.1%) had CSF cultures done with only four (0.07%) sample positive for bacteria (1 *Bacillus* species, 1 *Staphylococcus epidermidis*, 1 *Staphylococcus hominis*, and 1 unspecified bacteria positive). No cases of *Streptococcus pneumoniae* were identified on culture.

During hospitalization, 119 (2.0%) had a lumbar puncture and 97 (1.6%) had CSF chemistry and culture done. Four (0.08%) samples were positive for bacteria including *Escherichia coli*, *Micrococcus*, *Staphylococcus epidermidis* and *Streptococcus salivarius*, *Streptococcus mitis/oralis* as well as *Enterovirus* in two cases. No cases of *Streptococcus pneumoniae* were identified on culture. A total of 603 (9.9%) had a chest radiograph during their inpatient stay. Inpatient chest radiography findings showed that 354 (5.8%) had pneumonia, 48 (0.8%) had bronchitis, 2 (0.03%) had bronchiolitis, 3 (0.05%) had bronchiectasis, and 53 (0.9%) had other non-pneumonia abnormalities. In total, 698 (11.5%) of the study participants had been diagnosed with pneumonia on chest radiography at some point during hospitalization. There were two (0.03%) deaths recorded in the emergency room. One of the deaths was attributed to encephalitis and sepsis following infection. The cause of death in the second patient was related to serious underlying cardiac disease in a young infant.

Using the ViVI Score for cross-cohort comparison ::: Results:

The ViVI Disease Severity Scores showed a normal distribution with a mean score of 14.5 (SD: 6.0; range 0–34) at initial assessment (Figure 1). The ViVI Disease Severity Score was significantly higher in patients with the need for hospitalization (mean difference [95% CI]: –7.51 [–7.76 to –7.26];  $p < 0.001$ ), with a need for critical care facilities (mean difference [95% CI]: –6.24 [–6.58 to –5.91];  $p < 0.001$ ) as well as in those with signs of primary or secondary bacterial lower respiratory tract infections (mean difference [95% CI]: –6.46 [–6.71 to –6.20];  $p < 0.001$ ). The median Risk Factor Score in this cohort was 1 (IQR: 0–1); the median RF score was 0.88 (SD: 0.78) and scores ranged from 0 to 6.

Seasonality of respiratory viral infections ::: Results:

The CP analysis was applied to detect seasonal patterns for each virus detected in the QM Cohort. We define a virus to be seasonal if it is not present during the whole year. With this definition, we found that influenza viruses (Figure 2), as well as RSV and HMPV, showed a strong seasonal behavior (Figure 3) with predominance during the Northern Hemisphere winter months. ADV and HRV were ‘rapid cyclers’ with frequent and brief peaks throughout the year (Figure 3). The CP method showed that in a hospital-based syndromic surveillance system, seasons can be detected and defined in real time for each of the respiratory viruses. During the post-pandemic 2010/11 season for example, influenza A(H1N1)pdm09 viruses continued to predominate in the QM cohort. Influenza A(H3N2) viruses, on the other hand, were absent during the 2009/10 and 2010/11 seasons but replaced pandemic H1N1 strains during the subsequent season (see Figure 2). Also, differentiation of influenza B lineages revealed that Influenza B Yamagata and Victoria viruses did not always circulate annually but instead showed alternating patterns.

#### Seasonality of disease severity ::: Results:

The CP analysis was also used to identify fluctuations in average disease severity per calendar week in the QM program (Figure 4). The initial period until summer of 2011, when the QM program was restricted to once-weekly screening of in- and outpatients in the ED, is visually separated from the full surveillance phase beginning with the 2011/12 winter season, when daily screenings of all inpatients hospitalized with suspected ILI were added. The use of ViVI Disease Severity Scores during perennial, hospital-based surveillance provided standardized disease severity reports throughout the course of the year.

#### Disease severity with different respiratory viral (co)infections ::: Results:

There was a small but significant difference in ViVI Disease Severity Scores between those subjects where no viral etiology could be detected (mean ViVI Disease Severity Score: 13.85; SD 5.81), those identified with a single viral infection (mean ViVI Disease Severity Score: 14.90; SD 6.00) and those with more than 1 concurrent viral infection (mean ViVI Disease Severity Score: 15.73; SD 6.10);  $p$  (ANOVA) < 0.001. For each patient in the QM Cohort, we computed the overall ViVI Disease Severity Scores as well as the component of the DSU and DSC symptom category, respectively (Table 3). Average disease severity with different respiratory viral infections revealed that RSV induced the highest level of disease severity followed by HMPV, influenza A(H3N2), and HRV infections. Disease severity with ADV, influenza B, and influenza A (H1N1)pdm09 viruses remained below average (Table 3). The ViVI Disease Severity Score distributions including viral coinfections are displayed in Figure 5.

#### Comparison between disease severity and the consultation index ::: Results:

To illustrate the comparison, we computed the average ViVI Disease Severity Score per calendar week and compared to the Consultation Index during the same week (Figure 6(A)). As expected, no significant correlation was observed between weekly ViVI Disease Severity Scores and the Consultation Index (Pearson's correlation coefficient  $r = 0.10$ ;  $p = 0.1309$ ) during corresponding weeks, indicating disease severity and case numbers are not linked.

To allow visual interpretation, we also plotted the weekly average ViVI Disease Severity Score (in the ED prior to October 2011 and in ED and inpatient units thereafter) against the time course of the respective seasonal viruses. The results are shown in Figure 6(b). The viruses circulating (represented in % of all QM patients tested: y axis) are shown in relationship to the average ViVI Disease Severity Score during the respective calendar week. Some viruses peaked simultaneously with the average disease severity but a cumulative effect was more common. The effect of viruses prevalent during the summer months was more pronounced when inpatients were included in the QM Program, thus including severe cases requiring hospital admission.

#### RFs influencing disease severity ::: Results:

The Pearson's correlation coefficient  $r$  was 0.1923 indicating a statistically significant but weak positive correlation between ViVI Disease Severity and Risk Factor Scores ( $p < 0.001$ ). The distribution of ViVI Disease Severity Scores by different RFs is illustrated in Table 4.

To evaluate which of the RFs as defined by WHO [35] (Textbox 2) had the highest impact on disease severity (i.e. ViVI Disease Severity Score), we performed regression analysis as follows:  $\text{ViVI Disease Severity Score} = w_1 \times \text{RFs } 1, w_2 \times \text{RF } 2, \dots, w_n \times \text{RF } n$ . Regression analysis revealed that there was no specific set of variables that could be used to model this relationship significantly well. The best subset of RF variables was 'RF 1: Infant <2 years of age,' 'RF 3: Cardiac condition,' 'RF 2: Pulmonary condition,' 'RF 6: Obesity,' and 'RF 4: Diabetes' together yielded a  $R^2$  goodness-of-fit of 0.06. Using all RF variables yielded a  $R^2$  of 0.07. To further explore the relationship between age and RF, we studied median and mean ViVI Disease Severity Score in infants in children below 5 years of age, and in children aged 6 years and above (Table 5).

We then performed Pearson correlation to test for a potential relationship between ViVI Disease Severity Score and patient age. The Pearson correlation revealed  $r = -0.073$ , suggesting that there is in fact no significant correlation between the ViVI Disease Severity Score and patient age.

Disease severity in patients with and without antiviral/antibiotic prescription ::: Results:

An increasing ViVI Disease Severity Score indicates increasing disease severity. The key aspect of the ViVI Disease Severity Score is that it provides data standardization across the full spectrum of severity as well as comparison within a cohort, and between different seasons or sites. In the future, this may allow the comparison of various treatment decisions in clinical trials and observational settings.

As described above, physicians in routine care were unaware of the results of ViVI Disease Severity Score assessments by QM staff and reversely, QM staff were unaware of treatment decisions when assessing a patient. Analysis of ViVI Disease Severity Score results revealed that disease severity in patients (with any virus) who were prescribed neuraminidase inhibitors was 19.18 (95% CI: 17.62–20.74) compared to 14.52 (95% CI: 14.37–14.67) in individuals who were not prescribed neuraminidase inhibitors at the time of presentation to the ED (mean difference [95% CI]:  $-4.66$  [ $-6.24$  to  $-3.09$ ];  $p < 0.001$ ). In patients with PCR-confirmed influenza infection, this difference upheld: The mean ViVI Disease Severity Score in patients where physicians had decided to prescribe neuraminidase inhibitors was 20.33 (95% CI: 11.22–29.45) compared to 12.87 (95% CI: 12.40–13.33) in patients without antiviral therapy. The mean difference was  $-7.47$  (95% CI:  $-12.28$  to  $-2.65$ );  $p = 0.0024$ .

Similarly, patients, who had been prescribed antibiotics in routine care, also showed higher ViVI Disease Severity Score 16.73 (95% CI: 16.47–16.99) compared to a mean score of 13.51 (95% CI: 13.33–13.69) in individuals without antibiotic prescription (mean difference [95% CI]:  $-3.22$  [ $-3.53$  to  $-2.91$ ];  $p < 0.001$ ).

Using the ViVI Disease Severity Score to follow individual patients over time ::: Results:

Considering the variability in disease presentations and courses of illness with influenza and other respiratory viral infections in children, the ViVI Disease Severity Score is not intended to be validated against future clinical events or outcomes. To assess whether the ViVI Disease Severity Score could be used to standardize consecutive follow-up visits in clinical trials, a total number of 216 QM patients with influenza diagnoses were followed longitudinally with virology (PCR) and disease severity assessments over time.

The overall Pearson Correlation between ViVI Disease Severity Score and virus load (using cycle threshold = CT values) over time was 0.501. A closer look at the correlation histogram (Figure 7) revealed three major subgroups: The largest group of 161 patients can be categorized as having a moderate to strong positive correlation ( $r \geq 0.5$ ) between disease severity and viral load over time; a second group of 35 patients showed a strong negative correlation ( $r \leq -0.5$ ). A third group of 20 patients showed a weak (positive or negative) correlation ( $-0.5 < r < 0.5$ ). Preliminary decision tree analysis of these groups suggested that a ViVI Disease Severity Score below 11 and the RF 'infant below 2 years of age' were connected to a negative correlation between virus load and disease severity.

## Discussion:

Respiratory infections are among the most common reasons for children to be admitted to pediatric hospitals. Hospital-based surveillance of respiratory viral infections is of great value to understand the full disease spectrum, from mild symptoms to serious presentations. Children are the most avid transmitters of respiratory viral infections, and the information gained from syndromic surveillance in children's hospitals can complement decentralized sentinel surveillance systems in a meaningful way [133]. With the advent of rapid diagnostics and mobile health applications, it has now become possible to monitor virological and clinical end points in real time [134–140].

Traditionally, disease activity is monitored based on epidemiological parameters such as ILI or ARI incidence, hospitalization rates, or mortality [141,142]. The Consultation Index was developed by the Robert Koch Institute and has proven to be a sophisticated epidemiological tool to assess background ARI activity at representative sentinel practices. Fluctuations in ARI activity in private practices represent a useful indicator of disease burden based on actual case numbers. Reporting of the number of cases, however, does not reveal information on disease severity with each individual case [143].

The ViVI Disease Severity Score aims to fill this gap. The ViVI Disease Severity Score is a 22-item weighed clinical composite score consisting of DSU items reflecting 'regular' ILI activity and DSC items indicating 'high-impact' clinical presentations in the target population [144]. The ViVI Disease Severity Score opens avenues to new individual patient data (IPD) analyses, for example to identify clinically relevant seasonal patterns of disease severity linked to different viral diagnoses confirmed in the same group of patients.

The ViVI Disease Severity Score also allows consistent measurements of disease severity when following individual patients over time, as would be the case in clinical trials [40]. Follow-up assessments are useful whenever standardized severity data need to be recaptured over time. When frequent 'snap shots' of disease severity are combined with virology data, it may be possible to generate a 'moving image' with interesting new applications in clinical research. The introduction of standardized disease severity scores will facilitate head-to-head comparisons and the 'meta-analyzability' of clinical trials and observational studies. Full compliance of the ViVI Disease Severity Score mobile application with Clinical Data Interchange Standards Consortium (CDISC) standards further expands data interoperability and compliance with reporting formats to regulatory agencies [145–150].

It is important to note the scope of the proposed disease severity measure. This expert review does not intend to raise expectations that a disease severity score could or should be used to predict future events or physician behavior. Instead, we introduce a simple 22-item weighted clinical composite score allowing the assessor to translate the current condition of the patient into a two-digit number, which allows comparison of one patient to another, regardless of the setting. To this end, the paper provides the descriptive account of how a standardized score can be utilized to assess the relationships observed between the score and various (independent) treatment and management decisions for readers to draw their own conclusions about how they may in turn use the ViVI Disease Severity Score in clinical practice or research.

This expert review also introduces Time Series Analysis with Change Point Detection as a mathematical model applied, for the first time, to determining the timing and seasonality of respiratory viruses circulating in a hospital and emergency room. While the observation that several viruses may circulate in a seasonal pattern is not new, the authors demonstrate that purely data-inherent definitions of seasonality could be an interesting addition to traditionally used methods. Comparisons of average ViVI Disease Severity Score in the hospital system with the Consultation Index in the same calendar week (i.e. simultaneous ARI consultations in sentinel practices) revealed that the two parameters are intrinsically different. A 'heavy' season with a high frequency of ARI consultations is not the same as a 'light' season with fewer but more severe cases. The individual assessments in the QM cohort detected fluctuations in disease severity at a time when increases in overall ARI incidence in the general population were not evident. Especially during atypical influenza seasons with unusually few or unusually severe cases, the monitoring of disease severity in addition to incidence rates will provide important complementary information.

Standardized disease severity assessments also enable the cross-cohort comparison of disease burden between different viral pathogens. It is not surprising that RSV was identified as a key contributor to disease severity in a tertiary children's hospital, followed by HMPV disease. A better understanding of the real-world impact of different respiratory viruses on child health will help in the prioritization of drug and vaccine development.

The development of the ViVI Disease Severity Score is based on a systematic review of the published literature. The review showed that severity assessments have been inconsistent. Four clinical management parameters were used commonly as indicators of disease severity: hospitalization, intensive care treatment, oxygen supplementation, and mechanical ventilation (both invasive or noninvasive). The availability of any such measure, however, is highly dependent on the setting. The ViVI Disease Severity Score therefore uses the 'need for hospitalization' or 'need for ICU admission' (as determined by the assessor) instead. If the assessor determines that a patient would benefit from any such measures, the item can be scored regardless of the availability of ICU or hospital beds at the respective time or location.

To ensure inter-rater consistency, the QM team was specifically trained to apply established WHO definitions and standard criteria for the assessment of each aspect of the ViVI Disease Severity Score. For example, fever was defined according to Marcy et al. [151] and acute lower respiratory tract infection as per Roth et al. [152]. For use in multicenter settings, the ViVI Disease Severity Score App will include help menus in the user interface to ensure that assessors are aware of the same criteria and age-appropriate values. Acknowledging that the content, structure, and quality of standardized data are of paramount importance, the development team worked closely with



the CDISC to ensure full compliance of terminologies and data elements with industry and regulatory guidance.

The literature review showed that grading severity is not the same as predicting severity. Especially in young children, disease severity will fluctuate over time, until the episode is resolved eventually. The course of illness may or may not be linear. The ViVI Disease Severity Score is designed to help the physician measure and monitor the situation ad hoc, or repeatedly over time, but not to predict the future of the patient. Several scores have been designed, not to measure severity ad hoc, but to predict the likelihood of fatal outcomes in the future as is the case with the respiratory index of severity in children [153], the Pediatric Index of Mortality Score (PIMS) [83], and the pediatric risk of mortality score (PRIMS) [84,85]. These latter two scores (PIMS and PRIMS) were specific to RSV infections [83] in infants [84,85]. The Kristjansson Clinical Respiratory Score for RSV Infections in Children [82] was designed to include children beyond the infant age group. The index of severity was studied in bocavirus infections in infants and children <5 years [130], as was the symptom score for coronavirus infections in children [132]. Very few scores were developed to measure disease severity regardless of the type of respiratory virus causing the disease. The Clinical Severity Score was used to monitor RSV, HRV, and HMPV infections in children <3 years [91,117,122]. The systematic literature review, updated in 2016, confirmed that ViVI Disease Severity Score was the only score covering all pediatric age groups and any respiratory virus encompassing any of the key parameters outlined in the published literature to date [18,30]. The ViVI Disease Severity Score was also the only composite score that has been validated in a prospective cohort of more than 6000 children and adolescents from 0 to 18 years, yielding a normal distribution.

Regular severity assessments over time can be combined with CP detection methodology to detect of significant changes in disease severity in cohorts. Hospital surveillance will thus become feasible in real time, as rapid-turnaround diagnostic tests are evolving [154–157]. The use of rapid diagnostic tests can then be targeted according to the local surveillance information.

Bioinformatics analyses and machine learning algorithms may provide new avenues for the identification of virus-specific seasonality patterns [158].

During past influenza seasons, differences in the composition of subtypes and disease severity have been significant. The linkage of simultaneous virus surveillance with point-of-care disease severity assessments will advance the understanding of local epidemiology. Local epidemiology is key to understanding the impact of different strains on different populations. In North America, influenza A H3N2 viruses reappeared 1 year sooner than in Europe, i.e. in 2010/11 [159] followed by an unusually 'light' season with few or late cases during the winter of 2011/12 [160]. Public health agencies in the UK reported a particularly 'severe' season in 2010/11 [161,162], whereas Australia reported increased rates of severe influenza disease in 2014 [163–165], similar to Mexico during 2013/14. Classically, seasons have been regarded as 'severe' when coinciding with high overall case numbers, hospitalization rates, or mortality [166,167]. In the future, it will be important to distinguish the impact of fluctuations in influenza (sub)types on disease severity in specific patient groups, based on IPD.

The use of standardized measures of severity may also be helpful in the study of medical decision-making and diagnostic algorithms. Physicians in routine care often report that their decision to order virus diagnostics is often dependent on a variety of factors such as levels of training, media attention [168,169], specific requests by patients or parents, 'typical' versus 'atypical' disease presentations, availability and cost of diagnostic tests, insurance status of the patient, time constraints, etc. [20]. The same applies to the decision to hospitalize a patient. It is safe to assume that testing and rates of hospitalization are not the same at the beginning, peak, and end of an influenza season. Standardized disease severity scores may allow hospitals to set objective thresholds for diagnostic testing or admission decisions, depending on local conditions and epidemiology.

When population-based indicators are used instead of individual clinical outcome parameters, considerable bias may be introduced due to differences in patient reporting, access to health care [170] as well as physician awareness and reimbursement [171] creating challenges in global surveillance systems [172,173]. Some surveillance programs use retrospective chart reviews and ICD coding. ICD codes, however, do not always distinguish between laboratory-confirmed cases and clinical diagnoses [174].

Interpersonal variability and the unpredictable nature of respiratory viral infections pose a challenge to surveillance and preparedness programs [175]. Influenza seasons in particular vary with respect to case numbers and disease severity attributable to various viral subtypes and population strata [67,68,113,127]. The prospective monitoring of disease severity associated with

laboratory-confirmed diagnoses will help to delineate vulnerable subpopulations expressing disease severity differently compared to the population average. Real-time surveillance of disease severity may provide public health stakeholders with crucial information to adjust the allocation of hospital beds and resources [52]. The introduction of IPD disease severity assessments in a hospital-based surveillance system facilitates the timely identification of abnormal patterns of disease severity, i.e. through network analysis [176] or during time periods when the overall ILI disease severity is different from previous seasons or the rest of the year. Importantly, fluctuations in disease severity measured by the ViVI Disease Severity Score are independent of incidence-based surveillance indices.

Traditional disease severity estimates have focused on extreme presentations such as mortality rates [177,178] or ICU admission [179] but were not designed to monitor the full spectrum of mild-to-severe disease presentations. Additional granularity will be required for clinical trials. When the ViVI Disease Severity Score was used to follow patients longitudinally, disease severity was measured consistently from the time of initial presentation until resolution of symptoms. The ViVI Disease Severity Score has also been used to measure of subtle changes in disease severity in ICU patients requiring organ replacement therapy [40]. Once standardized scores are used consistently, this will open the path to head-to-head comparisons of antivirals and vaccines and to prospective IPD meta-analyses.

It will be important to investigate the complex relationship between virus load and disease severity and expected outcomes, which would provide important clues for clinical trial design [40]. Patients showing atypical patterns of disease severity for example (such as a negative correlation between virus load and disease severity) may represent individuals where antivirals do not exert the desired effect. Additional analyses are underway to understand this relationship better.

Standardized disease severity measures will facilitate biomarkers studies and the identification of viral and host factors associated with severe outcomes [180]. A precision medicine approach would lead to individualized risk communication strategies to improve the acceptance of vaccines and antivirals where they are most effective. The low uptake in influenza vaccines in the QM Cohort indicates that significant numbers of symptomatic influenza cases might have been prevented through immunization [181].

The presented work has several limitations: The current experience with the ViVI Disease Severity Score is based on a single-center tertiary care setting. Additional decentralized studies will be needed to validate the ViVI Disease Severity Score in international settings and in private practice networks, where severity may be lower. Further studies are planned in adults and the elderly, including the development of a compatible score for patient-reported outcomes. It is possible that different populations yield different results, but standardization is the prerequisite to study any such difference. Mobile applications will be particularly useful in low-resource settings, where disease severity may be higher and decisions have to be taken instantly. Finally, the effect of antiviral treatment or vaccine prevention on disease severity could not be assessed due to a minimal use of neuraminidase inhibitors and influenza vaccines in the current setting [182]. The only conclusion that can be drawn is that physicians in this setting hardly ever used antivirals but were more likely to prescribe antibiotics if a patient appeared severely ill, as expressed in significantly higher ViVI Disease Severity Scores [30]. It will be interesting to study decision-making processes and the impact of different forms of medical interventions on disease severity in a variety of settings in the future.

#### Expert commentary:

At this point, the majority of sentinel surveillance systems are laboratory based yielding limited clinical information but important data with respect to the evolution of influenza viruses, subtypes, resistance, seasonality, and transmissibility. It would be of great benefit to monitor disease severity individually, along with regional and geographic differences in virus circulation, using standardized disease severity measurements such as the ViVI Disease Severity Score.

Our contributions are the following: (A) The design of a hospital-based surveillance program and a unique QM cohort of more than 6000 children, where an independent QM team monitored patients daily using standardized clinical assessments and virology at the National Reference Centre for Influenza and Other Respiratory Viruses. (B) A novel disease severity score (the ViVI Disease Severity Score) and mobile application to detect specific changes in IPD and the individual course of illness in pediatric clinical trials and observational settings.

The presented tools are in line with the priorities issued by regulatory agencies with regards to data standardization and the development of clinical outcome measures for the development of new antivirals. With composite disease severity scores, the focus will shift from virological to

clinical end points, and the impact of therapeutic interventions on the quality of disease presentations. Only the systematic unbiased and prospective assessment of all cases, whether mild or severe, throughout several seasons, will provide objective insight into the actual disease burden with influenza and other respiratory viruses.

#### Five-year view:

Mobile health technologies enable new precision medicine approaches not only in clinical trials but also in routine patient care. Individualized disease severity assessments in children with influenza and other respiratory viruses will allow the physician to communicate better with the parent or patient, providing the current status as a validated measure of disease severity compared to similar age and population strata. In patients receiving antiviral therapy, progress can be measured and communicated accordingly and again, individually.

Most importantly, with the availability of validated disease severity measures and standardized datasets, the physician will be able to determine which patients may be 'lagging behind' in their response to therapeutic interventions. A better understanding of the complex relationship between virus load and disease severity in children with different respiratory viruses will provide important clues for a personalized approach to antiviral therapy.

Biomarker analyses linked to standardized disease severity assessments will help to elucidate why some patients improve rapidly as soon as virus loads decline, whereas a smaller group of patients does not improve as expected. This latter subgroup of patients may benefit from different therapeutic approaches, for example immunomodulation. Precision medicine tools such as the ViVI Disease Severity Score mobile application will provide important tools for the objective evaluation of new antivirals for soon-to-be treatable respiratory viruses.

#### Key issues:

Regulatory agencies and public health stakeholders have repeatedly called for international consensus on disease severity measures in influenza and other respiratory viruses. This need has become imminent with the rapid development of new anti-infective therapies for respiratory viral infections in children and adults. The challenge of data standardization is greatest in infants and young children, who may present with subtle and atypical symptoms. Based on a systematic review of the literature we developed a 22-item composite clinical score (the ViVI Disease Severity Score) for the immediate measurement of disease severity with acute respiratory infections the point-of-care. The ViVI Disease Severity Score was made available as a web-user interface and mobile application for validation in a quality management program including more than 6000 children 0–18 years of age. Linking standardized disease severity scores with rapid diagnostics will allow the instantaneous monitoring of incidence rates of acute respiratory viral infections along with the severity of each case. With this comprehensive manuscript, we are providing insight into the future of observational studies and clinical trials of antivirals for soon-to-be-treatable acute respiratory diseases in children. The reader is guided through novel analytic approaches that have become possible through rigorously standardized individual patient-data (IPD) analyses of disease severity.