

Self-sampling for analysis of respiratory viruses in a large-scale epidemiological study in Sweden

A Plymoth (amelie.plymoth@ki.se)¹, M Rotzén-Östlund^{2,3}, B Zweyberg-Wirgart^{2,3}, C G Sundin⁴, A Ploner¹, O Nyrén¹, A Linde⁵

1. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
2. Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
3. Department of Clinical Microbiology, Karolinska University Hospital, Solna, Sweden
4. Department of Communicable Disease Control, Södermanland County, Mälarsjukhuset, Eskilstuna, Sweden
5. Public Health Agency of Sweden, Solna, Sweden

Citation style for this article:

Plymoth A, Rotzén-Östlund M, Zweyberg-Wirgart B, Sundin CG, Ploner A, Nyrén O, Linde A. Self-sampling for analysis of respiratory viruses in a large-scale epidemiological study in Sweden. *Euro Surveill.* 2015;20(11):pii=21063. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=21063>

Article submitted on 21 February 2014 / published on 19 March 2015

Viral diagnosis of respiratory tract infections has so far required sampling by health professionals, hampering large-scale epidemiological studies of virus-specific disease outcomes. As part of a population-based, prospective study of work-related risk factors for transmission of viral infections (SWEDE-I), we developed a scheme for self-sampling with nasal swabs. Random selection from the gainfully employed population of a medium-sized town in central Sweden resulted in a study cohort of 2,237 men and women aged 25 to 63 years. From September 2011 through May 2012, the cohort reported all instances of respiratory tract infection or gastroenteritis and participants concomitantly sent self-sampled nasal swabs for analysis using regular mail. Diagnosis of 14 viruses was performed. A total of 1,843 samples were received. The week-wise average delay between disease onset and arrival of the specimens at the laboratory varied between four and six days, and the corresponding median delay was between 3.5 and six days. In line with previous community-based studies, picorna- and coronaviruses dominated in specimens obtained from the self-sampling scheme. The results of self-sampling were contrasted to those from contemporaneous routine clinical sampling, on the same age group, in the adjacent Stockholm county. Although higher proportions of positive samples for respiratory syncytial virus and influenza were observed in the clinical sampling scheme, estimations of seasonality for influenza A and picornaviruses derived from both schemes were similar. Our findings show that nasal self-sampling is feasible in large-scale surveillance of respiratory infections and opens new prospects for population-based, virologically verified research on virus spread, burden of disease, and effects of environmental factors or interventions.

Introduction

Epidemiological studies of respiratory infections require laboratory confirmation of causative agents. Even a syndrome such as influenza-like illness (ILI), which is regarded as marker for influenza in

routine surveillance, needs viral diagnosis in a subset of patients [1-3].

Until recently, viral sampling from the respiratory tract demanded professional involvement. This made large-scale sampling in epidemiological studies exceedingly expensive. Now, self-sampling by lay study participants and shipment of nasal swabs via regular mail should be feasible [4-6], since the sensitivity of polymerase chain reaction (PCR) enables detection of infectious agents at low concentrations. Further, viral genetic material remains stable under many conditions despite loss of infectivity and multiplex PCR assays capable of simultaneously examining many viruses can enable a comprehensive overview of circulating respiratory viruses [7].

Knowledge about the spread of specific viruses in the community is fundamental for successful prevention of epidemics. In many countries, the burden and spread of infections in society may differ substantially from what is seen among patients seeking healthcare. Even for influenza, only a minority of cases may show up in healthcare. The majority stay at home, and these cases probably account for the most substantial burden of disease, the largest cost due to loss of productivity, and likely form the main basis for spread of disease.

As part of the study of work environment and disease epidemiology-infections (SWEDE-I), we developed a scheme for self-initiated respiratory self-sampling with nasal swabs in a cohort that constituted a representative sample of the workforce in Eskilstuna – a medium-sized industrial town in central Sweden. The objective was to demonstrate the feasibility of nasal self-sampling as part of the population-based surveillance of respiratory virus infections in the adult, working population. Here, we describe the logistics and results of PCR-based analyses of 14 viruses. The virology results from this self-sampling were contrasted to those obtained in contemporaneous routine clinical

specimens, received during the same time period, from the same age group in the adjacent Stockholm area.

Methods

Study of work environment and disease epidemiology-infections

The study of work environment and disease epidemiology-infections (SWEDE-I) was designed to: (i) identify work-related factors associated with the risk of common acute respiratory infections and viral gastroenteritis, both overall and by causative viral agent, in order to pave the way for preventive measures; (ii) provide empirical data on factors that affect the probability of transmission of common viral infections in various work-related settings, in order to improve the epidemic models needed for predictions and planning when major outbreaks are anticipated. SWEDE-I used, for the first time, a newly developed and extensively tested population-based system for infectious disease surveillance with an analytical epidemiological approach. As part of this study, a scheme for self-initiated respiratory self-sampling with nasal swabs was developed.

Setting

We strived to conduct the self-sampling study within a fairly small, circumscribed and stable population. By restricting to such a population, the proportion of the source population that was constituted by participants would be sufficiently high to allow estimations of the period-specific activity of each virus of interest, as an indirect indicator of the probability of becoming exposed, based on the observations in the studied sample. In a small community with high participation density it is also easier to keep participants alerted to their reporting commitment since information about the study tends to propagate by word-of-mouth. Eskilstuna, a town with 97,373 inhabitants 110 km west of Stockholm, Sweden, corresponded well with our specifications and was chosen as the study site. Trade and industry is varied and includes several traditional manufacturing industries. Most gainfully employed people both live and work in the commune.

Study population

Gainfully employed people, aged 25–63 years, residing in Eskilstuna, constituted the source population. The sampling frame was provided by Statistics Sweden through cross-linkage of the continuously updated population register with the Employment Register. To achieve a premeditated sample size of 2,200, postal invitations were sent to an age- and sex-stratified random sample of 14,008 individuals. The expected under-representation of men and of the age stratum 25–44 years was compensated for in the recruitment by an over-sampling based on observed participation rates in earlier, similar population-based infectious disease surveillance studies [8].

The cohort filled in a series of web- or paper-based questionnaires which comprehensively probed into the

physical and psychological work environment, work tasks, contact patterns, and commuting behaviour, as well as into potential confounding factors such as diet, family structure, living conditions, medical history, personal characteristics, and physical activity.

Participant-initiated disease reporting and self-sampling of nasal secretion

Participants were instructed to self-report all onsets of fever ($>38^{\circ}\text{C}$), upper respiratory tract infection, and gastroenteritis, alone or in combination, immediately as they occurred during the entire study period from 1 September 2011 up to 31 May 2012. Reporting could be done via Internet or via telephone, using interactive voice response. When reporting an infection, the participants answered questions about symptoms in an automated, tree-structured interview. Based on pre-defined algorithms, the diseases could be classified as common cold [9], gastroenteritis, ILI [10], or other/unclassifiable. Frequent reminders by email and mail, and monthly newsletters reminded the participants of their commitment. Additionally, participants were requested to sample nose secretions concurrently with every symptom report. Two kits with nylon flocked dry swabs in plastic tubes (Copan Diagnostics, Inc., Murrieta, CA, US) and an instruction leaflet had been distributed to each participant shortly after entry into the cohort. Each kit was uniquely linked to the participant by a barcode label on the tube. The participant sent the sampled material to the Virology department, Clinical Microbiology Laboratory at Karolinska University Hospital, Solna, via regular, pre-paid mail. When a participant's last kit had been returned, a new one was supplied. The samples were stored at -70°C until tested for enterovirus, human coronavirus (hCoV) 229E, HKU1, NL63 and OC43, influenza A, A(H1N1) pdm09 and B, metapneumovirus (MPV), parainfluenza 1, 2 and 3, respiratory syncytial virus (RSV), rhinovirus, using in-house real-time PCR in 96-well plates [7]. Remaining material was stored in a biobank for five years.

To get confidential feedback on test results, each participant received an individually unique six-character code, which, combined with the unique national registration number, gave access to a secure webpage listing the participant's results by arrival date of the specimen. Each test result was accompanied by a text describing the virus and its associated disease.

Comparison with viral diagnoses in routine healthcare

The virological laboratory used for SWEDE-I also provides diagnostic services to the entire Stockholm county (population 2.1 million). For reference, aggregated weekly test results from all samples collected during the same time period (September 2011 to May 2012) and from persons in the same age group (25–63 years) as the SWEDE-I cohort, who had been diagnosed for some or all of the same viruses were extracted. These samples ($n=1,516$) represent a mixture of in- and

out-patients that were analysed for clinical reasons. Testing was administered in form of two standardised test packages: 73% of subjects (n=1,113) were only tested for influenza A (including A(H1N1)pdm09), influenza B and RSV. 27% (n=403), and typically in-patients, were tested for the full range of viruses analysed for in SWEDE-I, though without distinguishing between different picornaviruses.

Statistical methods

For each virus category, the total number and the percentage of positive swabs for each respiratory virus over all samples tested made are reported separately for the SWEDE-I material and the clinical samples. Percentages are given with exact 95% confidence intervals (CI) [11]. Hypothesis testing of equal percentages of positive tests for specific viruses in both materials was performed with Fisher's exact test (F-test).

To explore temporal trends in incidence, the proportion of new cases each week (as percentage of all cases during the nine-month study period) was computed for the four most frequent virus diagnoses (corona-, influenza A-, metapneumo- and picorna- viruses). Week-wise hypothesis testing of equal proportions in the SWEDE-I and clinical materials was done with F-test.

The distribution of sex, age, country of birth, and immigrant status was examined for SWEDE-I participants found to have a positive test for coronavirus, influenza A and picornavirus. Testing of the hypothesis that these distributions were the same for the virus-affected groups as for all participants who returned at least one nasal swab was done using chi-squared tests.

The weekly averages of number of days between reported onset of a disease episode and receipt of the corresponding nasal swab at the laboratory were plotted against study week, together with a loess smoothing curve for the mean, inversely weighted by weekly standard errors [12]. The smoothing curve was then tested against a null model of constant average delay using an approximate F-test [13]. Differences in this delay between individual positive and negative tests over the entire study period were tested using a Wilcoxon test. For all tests, a p-value of less than 0.05 was considered statistically significant.

Ethics

The study protocol was approved by the Stockholm Regional Ethics Review Board (dnr 2011/360–31/2). All participants gave their informed consent.

Role of the funding source

This research was funded in full by AFA Insurance, Stockholm, Sweden. The sponsor had no role in the conception, design, planning, execution, analysis, interpretation or publication of the study. The corresponding author had full access to all the data in the study.

Results

After two reminders, 2,237 of 14,008 invitees agreed to participate in the SWEDE-I cohort (participation rate 16%). Some key characteristics of the cohort are exhibited in Table 1. The participants sent in 1,843 nasal swabs and made 2,119 disease reports, giving a sampling rate of 87%. Of the nasal swabs, 876 (47.5%; 95% CI: 45.3–49.8%) were shown to contain at least one

TABLE 1

Demographic and clinical characteristics of the cohort for the study of work-related risk factors for transmission of viral infections (SWEDE-I), Sweden, September 2011–May 2012 (n=2,237 participants)

Characteristic	n/N ^{a,b} (%)
Sex	
Men	918/2,237 (41)
Women	1,319/2,237 (59)
Age in years	
Mean (standard deviation)	47 (10)
Median (interquartile range)	46 (39–56)
Age group in years	
25–34	284/2,237 (13)
35–44	718/2,237 (32)
45–54	592/2,237 (26)
55–63	643/2,237 (29)
Household size^c	
1 person	251/1,815 (14)
2 persons	705/1,815 (39)
3 persons	326/1,815 (18)
4 persons	396/1,815 (22)
≥5 persons	137/1,815 (8)
Children below 13 years-old in household	
0 children	1,248/1,815 (69)
1 child	245/1,815 (13)
2 children	235/1,815 (13)
≥3 children	87/1,815 (5)
Highest attained education	
Secondary school (≤9 years)	160/1,798 (9)
Sixth Form (11–13 years)	596/1,798 (33)
University/college < 3 years	260/1,798 (15)
University/college ≥ 3 years	596/1,798 (33)
Other post-sixth form education	186/1,798 (10)
Healthcare work	
Yes	349/1,906 (18)
No	1,557/1,906 (82)
Number of reported disease events	
0	1,158/2,237 (52)
1	654/2,237 (29)
2	267/2,237 (12)
≥3	158/2,237 (7)

^a A total 2,237 participants were in the cohort but not all responded to all questions asked in the questionnaire.

^b Unless otherwise specified.

^c Including the index participant.

virus (henceforth referred to as ‘positive tests’). Since 21 of the samples each contained two or more different viruses (two viral diagnoses in 20 samples, three in one), the total number of virus diagnoses was 898.

The number of returned swabs peaked in the last week of September 2011 (week 39) and fell until the last week of November (week 48), when a phase with variable inflow ensued (Figure 1A). After the last week of February 2012 (week 9), the numbers decreased until the end of the study. The crude number of positive tests showed a similar pattern, but there was a more distinct upward tendency from late November, a climax in mid-February, and a gradual decrease until late April. In the beginning of the self-sampling study, the proportion of positive tests remained at ca 40%, but from mid-November the proportion increased, until it exceeded 60% in early April (week 14) (Figure 1B). Then it abruptly fell back to around 45%.

With the exception of the first and last two weeks of the study, the week-wise average delay between onset of disease episodes and arrival of the specimens at the laboratory varied between four and six days, and the corresponding median delay was between 3.5 and six days. Figure 2 suggests that this average increased slightly towards the end of the study period, although the trend indicated by the smoothing curve failed to reach formal significance ($p=0.06$). Overall, the delay between disease onset and sample arrival appeared fairly stable. Negative tests were slightly but significantly skewed towards longer delays ($p=0.004$).

Of 1,212 episodes with reported nasal discharge, 679 (56.0%) showed positive tests, in stark contrast to 29 (12.1%) of 239 episodes without nasal discharge. Test-negative episodes without nasal discharge were evenly distributed across the entire study period (data not shown).

FIGURE 1

Weekly number of swabs received by the cohort of the study of work-related risk factors for transmission of viral infections (SWEDE-I) and proportion of positive tests among all swabs received, Sweden, September 2011–May 2012 (n=1,843 swabs)

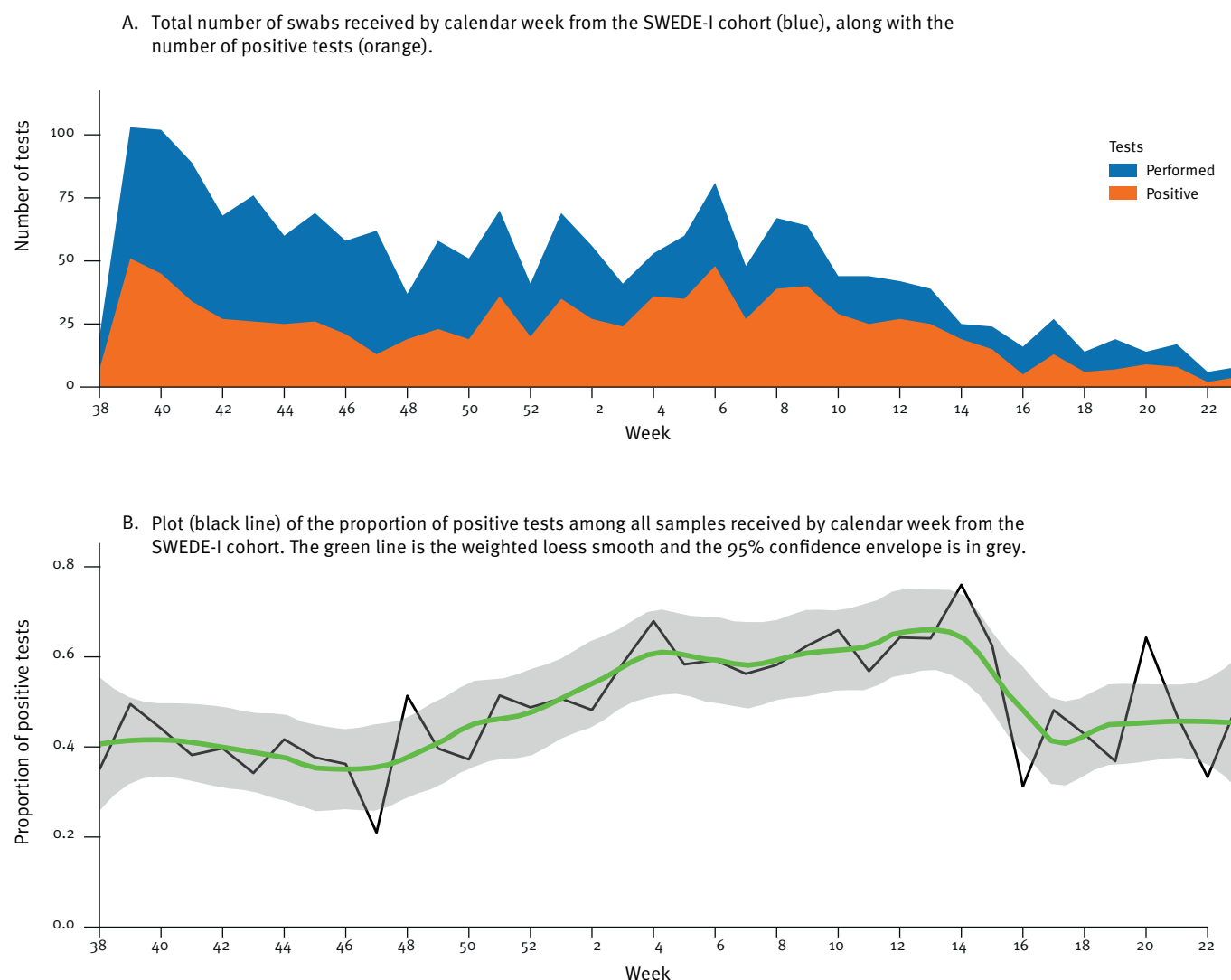
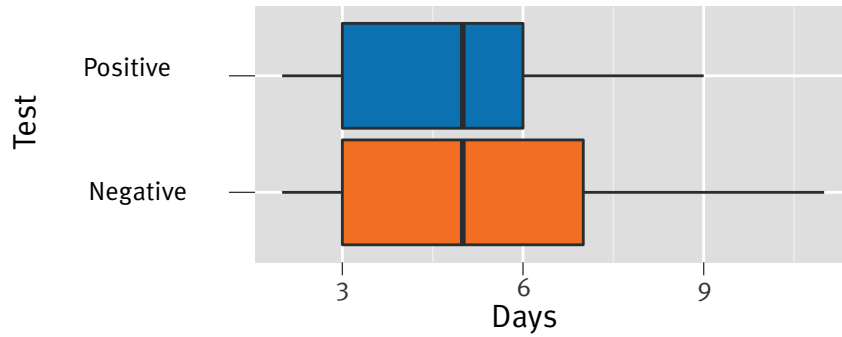


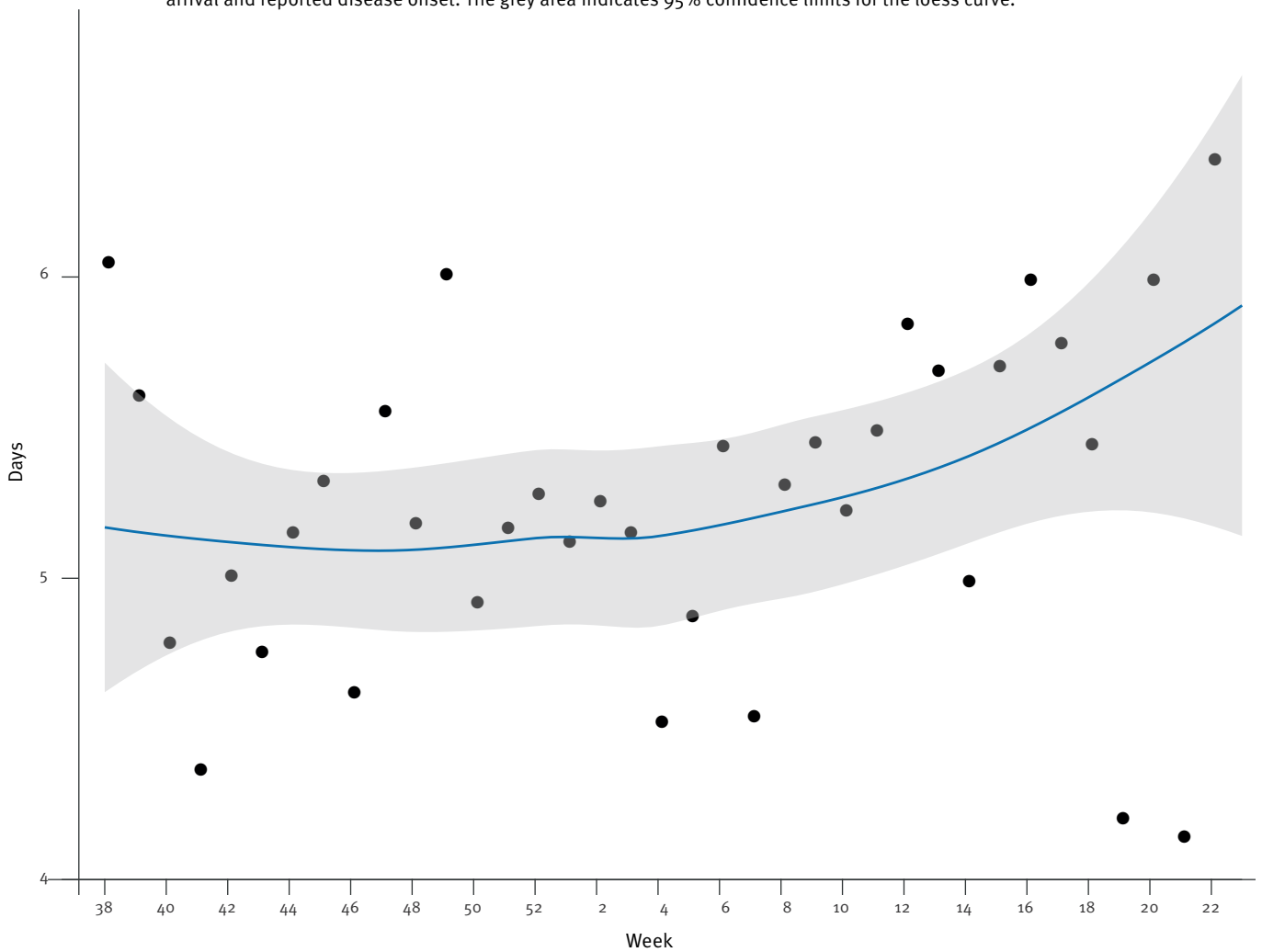
FIGURE 2

Delay between specimen arrival at the laboratory and reported disease episode onset in the cohort performing self-sampling, study of work-related risk factors for transmission of viral infections (SWEDE-I), Sweden, September 2011–May 2012 (n=876 swabs)

A. Distribution of number of days between reported disease onset and arrival of the specimen at the laboratory for positive and negative tests; the modified boxplots show quartiles and median (box) as well as 5% and 95% quantiles (whiskers).



B. Loess smoothing curve (blue) weighted by standard errors of weekly averages of number of days elapsed between specimen arrival and reported disease onset. The grey area indicates 95% confidence limits for the loess curve.



Pattern of virus-specific diagnoses

Percentages of virus-specific diagnoses among all samples tested in the SWEDE-I cohort are listed in Table 2 (column 2). In the SWEDE-I material, rhinoviruses were the most common of all tested viruses (20.8% of all samples) and dominated the picornavirus group. Coronaviruses, dominated by HKU1, were found in 16.2% of the samples, followed by seasonal influenza A in 4.6% of the samples and MPV in 2.9%. Among test-positive samples from patients without nasal discharge, the distribution of virus types was essentially the same as that in the entire SWEDE-I cohort (data not shown).

Columns 4 of Table 2 display the diagnostic yield from the clinical material. As for the SWEDE-I samples, corona-, influenza A and picorna-, viruses dominated, in the clinically-isolated samples, however, the rank order differed substantially and the proportion of influenza and RSV positive samples was significantly higher among such samples ($p \leq 6e-05$).

Seasonality of virus-specific diagnoses

The seasonal distributions of virus-specific diagnoses across the study period are shown in Figure 3 as weekly proportions of all specific diagnoses in the study period. From the SWEDE-I samples we found that picornaviruses, which were dominated by rhinoviruses occurred during the entire study period, but with a distinct peak in the last week of September (week 39). The season for coronaviruses lasted from early November until early May (week 18), with a climax in the second week of February (week 6). Seasonal influenza A peaked during the first three weeks of March (weeks 9–11). MPV occurred with three distinct peaks four to seven weeks apart between late December (week 52) and late March (week 13). Again the seasonal pattern among test-positive samples from patients without nasal discharge was very similar to the whole SWEDE-I cohort as just described (data not shown).

Figure 3 also shows the corresponding proportions of virus-specific diagnoses in the clinically-isolated

TABLE 2

Numbers, and percentages among all samples tested, of positive diagnoses for respiratory viruses found respectively for self-sampled and clinically-sampled swabs, Sweden, September 2011–May 2012

Test result	SWEDE-I self-sampled swabs		Clinically sampled swabs		P-value
	Number of positive samples among all samples tested n/N (%) ^a	95% CI	Number of positive samples among all samples tested n/N (%) ^{b,c}	95% CI	
Picornavirus	416/1,843 (22.6)	20.7–24.6	44/403 (10.9)	8.0–14.4	4e-08
Rhinovirus	383/1,843 (20.8)	18.9–22.7	NA	NA	NA
Enterovirus	18/1,843 (1.0)	0.6–1.5	NA	NA	NA
Undecided	15/1,843 (0.8)	0.5–1.3	NA	NA	NA
Coronavirus	298/1,843 (16.2)	14.5–17.9	22/403 (5.5)	3.5–8.1	1e-09
HKU1	167/1,843 (9.1)	7.8–10.5	9/403 (2.2)	1.0–4.2	3e-07
229E	85/1,843 (4.6)	3.7–5.7	7/403 (1.7)	0.7–3.5	0.008
OC43	41/1,843 (2.2)	1.6–3.0	4/403 (1.0)	0.3–2.5	0.120
NL63	5/1,843 (0.3)	0.1–0.6	2/403 (0.5)	0.1–1.8	0.616
Influenza virus	88/1,843 (4.8)	3.9–5.8	380/1,516 (25.1)	22.9–27.3	<2e-16
A seasonal	84/1,843 (4.6)	3.7–5.6	349/1,516 (23.0)	20.9–25.2	<2e-16
A(H1N1)pdm09	0/1,843 (0.0)	0.0–0.2	13/1,516 (0.9)	0.5–1.5	3e-05
B	4/1,843 (0.2)	0.1–0.6	18/1,516 (1.2)	0.7–1.9	8e-04
Metapneumovirus	54/1,843 (2.9)	2.2–3.8	22/403 (5.5)	3.5–8.1	0.015
Parainfluenza virus (PIV)	25/1,843 (1.4)	0.9–2.0	9/403 (2.2)	1.0–4.2	0.182
PIV1	14/1,843 (0.8)	0.4–1.3	2/403 (0.5)	0.1–1.8	0.752
PIV3	10/1,843 (0.5)	0.3–1.0	7/403 (1.7)	0.7–3.5	0.021
PIV2	1/1,843 (0.1)	0.0–0.3	0/403 (0.0)	0.0–0.9	1.000
Respiratory syncytial virus (RSV)	17/1,843 (0.9)	0.5–1.5	42/1,516 (2.8)	2.0–3.7	6e-05

CI: confidence interval; NA: not applicable; SWEDE-I: study of work-related risk factors for transmission of viral infections. In the SWEDE-I study, 855 samples generated one viral diagnosis, 20 two diagnoses, and one three diagnoses.

^a The rate of positives for the influenza viruses and RSV in the SWEDE-I samples was 105/1,843 = 5.7% and for the other viruses it was 793/1,843 = 43.0%. The overall positivity rate was 876/1,843 = 47.5%.

^b The rate of positives for influenza viruses and RSV in the clinical samples was 422/1,516 = 27.8% and for the other viruses, which were only tested for in a subset of patients, it was 97/403 = 24.1%. The overall positivity rate cannot be calculated for the clinical material.

^c Of 1,515 total clinical samples, only 403 (27%), typically from in-patients, were tested for the full range of viruses analysed for in SWEDE-I, though without distinguishing between different picornaviruses. All samples were however tested for influenza A (including A(H1N1)pdm09), influenza B and RSV.

samples. Despite the difference in absolute frequencies seen in Table 2, the resulting proportions obtained respectively within the self-sampling and clinical-sampling schemes indicated seasonal occurrences that were overall similar. Influenza A tracked extremely well between the two schemes, the only difference ($p=0.01$) being a second dominant peak in the SWEDE-I cohort, two weeks after the common peak. For the corona-, metapneumo- and picornaviruses, significant differences in weekly proportions of positive swabs among the two sampling schemes were mostly observed towards the end of the study period, with the exception of one obvious peak in week 11 for coronavirus in

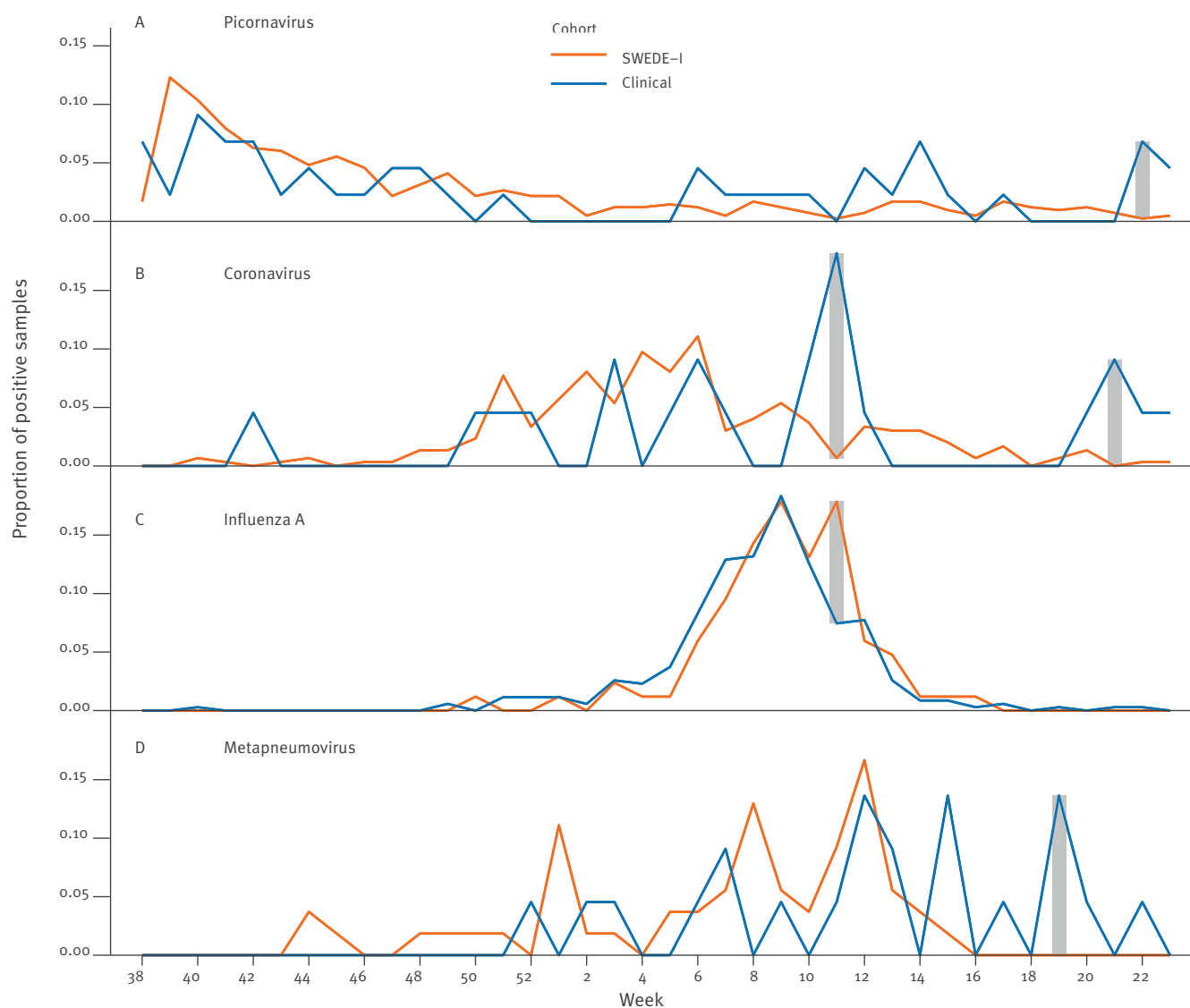
the clinical material, which was completely absent in the SWEDE-I material ($p=0.0004$). The counts for the other viruses considered in this study were too few to allow for seasonal analyses.

Common infections and demographics

The distributions of age, sex and foreign background were remarkably similar among SWEDE-I participants with positive tests for, respectively, coronavirus, influenza A and picornavirus (Table 3). None of these virus-positive groups differed significantly from the total group with tested nasal swabs. People born abroad or with an immigrant background were similarly

FIGURE 3

Weekly proportions of samples positive for (A) picornavirus, (B) coronavirus, (C) influenza A and (D) metapneumovirus relative to the total positive respective samples during the whole study obtained in the self-sampled and clinically-sampled materials, Sweden, September 2011–May 2012



SWEDE-I: study of work-related risk factors for transmission of viral infections.

Each data point represents the proportion of positive samples for a given virus that was received during the specified week relative to the season's total number of positive samples for this virus. Red lines represent self-sampled material from the SWEDE-I cohort in Eskilstuna, blue lines represent samples retrieved from contemporaneous routine clinical material from patients of the same age group seen at Karolinska University Hospital in Stockholm. Weeks in which the proportions differed significantly ($p < 0.05$) between the two materials are marked with a grey bar.

represented in the virus-positive groups. However, the age distributions among participants with coronavirus- and influenza A-positive infections differed significantly; while the majority of the former were 35–54 years of age, influenza A affected mainly participants in the youngest (25–34 years) and oldest (55–63 years) age groups ($p=0.03$).

Discussion

This study demonstrates the feasibility of nasal self-sampling as part of population-based surveillance of respiratory virus infections. During one season 1,843 samples, corresponding to 1.1 per person-year, could be evaluated. Previous reports have indicated that self-sampling earlier in the disease likely compensates for possible losses in sample quality [14,15]. The week-wise average delay between onset of disease episodes and arrival of the specimens at the laboratory varied between four and six days, and the corresponding median delay was between 3.5 and six days. This delay is considered acceptable in terms of sample quality.

While we lack a formal validation against a gold-standard method, we argue for the validity of our results based on a number of separate lines of evidence. With regard to self-reporting of ILI/ acute respiratory infection (ARI), the framework for self-initiated, event-driven infectious disease reporting that we employed had already been developed for a Swedish population-based cohort [8] and used for population-based surveillance in Stockholm County since 2007 [16]. A separate validation study concluded that while there

was significant under-reporting of disease (estimated at 60%), this level of under-reporting was remarkably constant over time and across seasons [16], so that a simple constant correction factor can potentially restore validity of incidence rates, at least in terms of reported disease incidence.

It is possible that the additional requirement of collecting and mailing a nasal sample may have led to increased study fatigue and correspondingly increased under-reporting in the SWEDE-I study, compared with the previous studies: in Figure 1A, we see indeed that the number of tests performed decreases from more than 100 samples/week at the very beginning to ca 20 samples/week at the end. However, Figure 1B indicates that the proportion of positive samples was at about the same level of ca 40% at both times, with in between a peak of ca 60% positive samples coinciding with the peak of the influenza A season seen in Figure 3. Also, within the study period, the numbers of self-sampled specimens submitted to the laboratory decreased more after week 9. This is in agreement with syndromic surveillance in adults for the same period in the whole of Sweden, based on calls for fever and cough to a medical advice line as an indicator for respiratory infections, which shows a sharp decline of contacts from week 9, the last week in February [17]. Taken together, it appears that the fluctuation in positive samples at least is driven more by seasonal and disease-related factors than by varying levels of participation in the study.

TABLE 3

Comparison according to selected demographic variables of SWEDE-I participants testing positive for coronavirus, influenza A, or picornavirus with all participants who returned a nasal swab, Sweden, September 2011–May 2012 ($n=1,843$)

Demographic characteristics	All incoming nasal swabs ($n=1,843$)	Picornavirus positive ($n=416$)	Coronavirus positive ($n=292$)	Influenza A virus positive ($n=84$)
	n (%)	n (%)	n (%)	n (%)
Sex				
Men	608 (33)	142 (34)	104 (36)	24 (29)
Women	1,235 (67)	274 (66)	188 (65)	60 (71)
Age group in years				
25–34	256 (14)	55 (13)	30 (10)	15 (18)
35–44	618 (33)	140 (34)	90 (31)	24 (29)
45–54	436 (24)	108 (26)	83 (28)	13 (15)
55–63	533 (29)	113 (27)	89 (31)	32 (38)
Foreign born				
Yes	103 (5)	21 (5)	15 (5)	5 (6)
No	1,305 (71)	297 (71)	203 (70)	63 (75)
Missing	435 (24)	98 (24)	74 (25)	16 (19)
Immigrant status^a				
Yes	304 (16)	60 (14)	48 (16)	14 (17)
No	1,102 (60)	257 (62)	170 (58)	54 (64)
Missing	437 (24)	99 (24)	74 (25)	16 (19)

SWEDE-I: study of work-related risk factors for transmission of viral infections.

^a At least one parent born outside Sweden.

With regard to the distribution and burden of viruses in SWEDE-I, we found a similar pattern of rhinovirus-coronavirus-influenza as the three most common diagnoses (at 23%/ 16%/ 5%, respectively) seen in community-based studies in England in the 1990s (34%/ 14%/ 9% for a population aged 0–60+ years [18] and 52%/ 26%/ 10% for a population aged 60–90 years [19]). These community-based studies used active follow-up, with sampling by health professionals, and diagnosis through a combination of virus isolation and serology.

This is strikingly different from the concurrent clinically-isolated sampling scheme results, where the proportion of influenza-positive samples (25%) dominated all other infections, with seasonal influenza A, A(H1N1)pdm09, B being significantly more frequent than in SWEDE-I. At the same time, RSV and MPV were also significantly more common in the clinical materials. Although the lower frequency of viruses causing severe infections in SWEDE-I could be partly due to the self-sampling method, it seems most likely that patients infected with these viruses are overrepresented in healthcare.

Even though we have not been able to demonstrate conclusively in this study that self-sampling has the same sensitivity as healthcare based sampling, our results strongly support the use of the SWEDE-I methodology for influenza surveillance.

With regard to the timing of the circulation of different viruses, we found that the seasonality patterns obtained were rather similar between the SWEDE-I and clinical schemes. This confirms that clinical identification parallels societal spread, as seen in similar self-sampling studies previously [15], but is by no means a measure of societal spread intensity. The higher proportion of positive tests in the SWEDE-I cohort was explained by the abundance of picorna- and coronavirus infections. Self-sampling earlier in the disease may have contributed to their frequent detection, but the most apparent explanation for their scarcity in the clinical material is that they are rarely direct causes of severe disease among adults [20,21]. Interestingly, when influenza A peaked in the SWEDE-I cohort, only ca 30% of the samples were positive for influenza, and another 30% were positive for other tested viruses. This underlines the importance of virological testing to verify that acute respiratory disease is caused by influenza also during the epidemic period.

In the SWEDE-I cohort, the proportion of positives for picorna virus, corona virus and influenza viruses in adult persons from various demographic groups was very similar to the proportion among all samples obtained (Table 3). The similarity in this non-healthcare selected, adult population is obvious both when age, sex and ethnicity are considered. It is difficult to make any other interpretation than that spread of these viruses, with accompanying respiratory symptoms, is rather homogenous among adults of similar age in the

society. The low rate of positives for the other viruses prevented a similar analysis.

Noticeable limitations of the study include the absence of formal validation against gold standard testing, uncertain external validity due to low participation in the invited representative sample, and probable under-reporting among participants. Men, young age groups, and low-educated people were somewhat under-represented in a similar cohort [8]. The number of disease reports per person-year in this study is very similar to a previous validation study in similar cohorts [16]. In the previous study, a relatively constant under-reporting of 60% was identified, based on random control questionnaires on health status the previous week. Assuming a constant overall incidence of virus infections from year to year, the under-reporting was likely similar in the present study. Clearly, more research is needed to improve the completeness of disease reporting. Additional reminders and other incentives may be required.

The rate of positivity was further considerably higher among individuals whose disease was associated with nasal discharge than among those without. We found no indications that this disfavoured any specific virus, but further research is needed to verify whether patients without rhinitis are virally infected, and if so, to improve sampling.

This is a large-scale epidemiological study where self-reporting, self-sampling and modern PCR-based diagnosis were combined for investigation of virus-specific respiratory infection incidence on the population level. The logistics around reporting and self-sampling functioned exceptionally well. Of major importance was the sensitivity of the virological assays used. The methodology has been evaluated [7] and the sensitivity appears to be optimal. The participants received written instructions on how to perform the self-sampling and the instructions were also available on the study web page. In addition the participants could also call the staff at the study centre to ask questions. Shortly after having sent in a sample the participant could log in with their unique code at the study web page, for access to a secure website with their viral test results. The major cost of a virological study is the laboratory analyses, and to contain costs, samples can be stored at -70 °C until analysis in batches during periods of low workload since analysis is not necessary for clinical purpose. While the present cost of virological analyses makes routine sampling for analyses of respiratory viruses in the population unjustifiable, the feasibility of large-scale self-sampling in epidemiological studies may importantly advance the understanding of burden of disease and factors affecting spread.

The discrepancies and similarities with findings in clinical specimens seem logical, and calculations for influenza result in a very relevant incidence for the included population. For some of the viruses, a laboratory

comparison of sensitivity for nasal vs nasopharyngeal aspirates is desirable, but the fact that self-sampling is performed very early during the disease may compensate for a higher sensitivity of clinical nasopharyngeal sampling.

This successful deployment of self-sampling is applicable everywhere and it can be extended to other groups than working adults, and to various geographical areas, so long as the mail transport is reasonably efficient. We believe it may be an important tool in further research on spread of viruses in the population and the effect of interventions such as vaccination. Self-sampling for vaginal and rectal material has already been introduced for diagnosis of venereal diseases [22]. This sampling method can certainly support clinical and syndromic surveillance, as previously suggested [15].

Acknowledgments

This research was funded in full by AFA Insurance, Stockholm, Sweden, Grant Dnr 090282. ON was supported by the European Union Seventh Framework Programme (grant number EPIWORK 231807) and the Strategic Research Program in Epidemiology at Karolinska Institutet.

Conflict of interest

None declared.

Authors' contributions

ON, APly, and AL designed the study; APly and ON were responsible for the enrolment of participants, organization of fieldwork, and data collection; AL and C-GS supported the fieldwork; MRÖ and BZW performed viral analyses and supplied data on viral diagnoses in routine clinical samples; APlo, MRÖ and APly did the statistical analysis; AL, MRÖ, ON, APly, APlo, BZW, and C-GS interpreted the data; APly, ON, APlo, MRÖ, BZW and AL wrote and revised the draft report; APly, ON, APlo, MRÖ, BZW, AL and C-GS reviewed and revised the report. All authors approved the final version of the report for submission.

References

1. Fleming DM. Weekly Returns Service of the Royal College of General Practitioners. *Commun Dis Public Health*. 1999;2(2):96-100. PMID:10402742
2. World Health Organization (WHO). WHO global influenza surveillance network: manual for the laboratory diagnosis and virological surveillance of influenza. Geneva: WHO Press; 2011.
3. European Influenza Surveillance Network (EISN). Stockholm: European Centre for Disease Control and Prevention. Available from: <http://ecdc.europa.eu/en/activities/surveillance/EISN/Pages/index.aspx>
4. Cooper DL, Smith GE, Chinemana F, Joseph C, Loveridge P, Sebastianpillai P, et al. Linking syndromic surveillance with virological self-sampling. *Epidemiol Infect*. 2008;136(2):222-4. <http://dx.doi.org/10.1017/S0950268807008412> PMID:17394678
5. Akmatov MK, Krebs S, Preusse M, Gatzemeier A, Frischmann U, Schughart K, et al. E-mail-based symptomatic surveillance combined with self-collection of nasal swabs: a new tool for acute respiratory infection epidemiology. *Int J Infect Dis*. 2011;15(11):e799-803. <http://dx.doi.org/10.1016/j.ijid.2011.07.005> PMID:21852171
6. Akmatov MK, Pessler F. Self-collected nasal swabs to detect infection and colonization: a useful tool for population-based epidemiological studies? *Int J Infect Dis*. 2011;15(9):e589-93. <http://dx.doi.org/10.1016/j.ijid.2011.04.009> PMID:21641847
7. Tiveljung-Lindell A, Rotzén-Ostlund M, Gupta S, Ullstrand R, Grillner L, Zweyberg-Wirgart B, et al. Development and implementation of a molecular diagnostic platform for daily rapid detection of 15 respiratory viruses. *J Med Virol*. 2009;81(1):167-75. <http://dx.doi.org/10.1002/jmv.21368> PMID:19031448
8. Bexelius C, Merk H, Sandin S, Nyrén O, Köhlmann-Berenzon S, Linde A, et al. Interactive voice response and web-based questionnaires for population-based infectious disease reporting. *Eur J Epidemiol*. 2010;25(10):693-702. <http://dx.doi.org/10.1007/s10654-010-9484-y> PMID:20596884
9. Gwaltney JM Jr. Rhinovirus colds: epidemiology, clinical characteristics and transmission. *Eur J Respir Dis Suppl*. 1983;128(Pt 1):336-9. PMID:6311603
10. European Commission. Commission implementing decision (2012/506/EU) of 8 August 2012 amending Decision 2002/253/EC laying down case definitions for reporting communicable diseases to the Community network under Decision No 2119/98/EC of the European Parliament and of the Council. *Official Journal of the European Union*. Luxembourg: Publications Office of the European Union. 27.9.2012. Available from: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:262:0001:0057:EN:PDF> 2012/506/EU
11. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med*. 1998;17(8):857-72. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19980430\)17:8<857::AID-SIM777>3.0.CO;2-E](http://dx.doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E) PMID:9595616
12. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc*. 1979;74(368):829-36. <http://dx.doi.org/10.1080/01621459.1979.10481038>
13. Fox J. Sectio 18.1.2. In: *Applied Regression Analysis and Generalized Linear Models*. 2nd ed. Sage; 2008. p. 492.
14. Ohrmalm L, Wong M, Rotzén-Ostlund M, Norbeck O, Broliden K, Tolfvenstam T. Flocked nasal swab versus nasopharyngeal aspirate for detection of respiratory tract viruses in immunocompromised adults: a matched comparative study. *BMC Infect Dis*. 2010;10(1):340. <http://dx.doi.org/10.1186/1471-2334-10-340> PMID:21110854
15. Elliot AJ, Powers C, Thornton A, Obi C, Hill C, Simms I, et al. Monitoring the emergence of community transmission of influenza A/H1N1 2009 in England: a cross sectional opportunistic survey of self sampled telephone callers to NHS Direct. *BMJ*. 2009;339(aug27 2):b3403. <http://dx.doi.org/10.1136/bmj.b3403> PMID:19713236
16. Merk H, Köhlmann-Berenzon S, Bexelius C, Sandin S, Litton JE, Linde A, et al. The validity of self-initiated, event-driven infectious disease reporting in general population cohorts. *PLoS One*. 2013;8(4):e61644. <http://dx.doi.org/10.1371/journal.pone.0061644> PMID:23613891
17. Smittskyddsinstitutet (SMI). Influenza in Sweden season 2012-2013. Solna: SMI; 23 October 2014. Available from: <http://www.folkhalsomyndigheten.se/pagefiles/12992/Influenza-in-Sweden-2012-2013.pdf>
18. Monto AS, Sullivan KM. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol Infect*. 1993;110(1):145-60. <http://dx.doi.org/10.1017/S0950268800050779> PMID:8432318
19. Nicholson KG, Kent J, Hammersley V, Cancio E. Acute viral infections of upper respiratory tract in elderly people living in the community: comparative, prospective, population based study of disease burden. *BMJ*. 1997;315(7115):1060-4. <http://dx.doi.org/10.1136/bmj.315.7115.1060> PMID:9366736
20. Mackay IM. Mackay IM. Human rhinoviruses: the cold wars resume. *J Clin Virol*. 2008;42(4):297-320. <http://dx.doi.org/10.1016/j.jcv.2008.04.002> PMID:18502684
21. van der Hoek L. Human coronaviruses: what do they cause? *Antivir Ther*. 2007;12(4 Pt B):651-8. Review. PMID:17944272.
22. Gaydos CA, Quinn TC. Urine nucleic acid amplification tests for the diagnosis of sexually transmitted infections in clinical practice. *Curr Opin Infect Dis*. 2005;18(1):55-66. <http://dx.doi.org/10.1097/00001432-200502000-00010> PMID:15647701