HAL open science

Monitoring sick leave data for early detection of influenza outbreaks

Tom Duchemin, Jonathan Bastard, Pearl Anne Ante-Testard, Rania Assab, Oumou Salama Daouda, Audrey Duval, Jérôme-Philippe Garsi, Radowan Lounissi, Narimane Nekkab, Helene Neynaud, et al.

To cite this version:

Tom Duchemin, Jonathan Bastard, Pearl Anne Ante-Testard, Rania Assab, Oumou Salama Daouda, et al.. Monitoring sick leave data for early detection of influenza outbreaks. BMC Infectious Diseases, 2021, 21, 10.1186/s12879-020-05754-5 . hal-03114532v2

HAL Id: hal-03114532

https://hal.science/hal-03114532v2

Submitted on 12 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Duchemin et al. BMC Infectious Diseases (2021) 21:52
https://doi.org/10.1186/s12879-020-05754-5

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

Monitoring sick leave data for early detection of influenza outbreaks

Tom Duchemin[1,2*], Jonathan Bastard[1,3,4,5], Pearl Anne Ante-Testard[1,4], Rania Assab[1], Oumou Salama Daouda[1], Audrey Duval[1,3,5,6], Jérôme-Philippe Garsi[1], Radowan Lounissi[2], Narimane Nekkab[1,7], Helene Neynaud[1], David R. M. Smith[1,3,5], William Dab[1], Kevin Jean[1,4], Laura Temime[1,4] and Mounia N. Hocine[1]

Abstract

Background: Workplace absenteeism increases significantly during influenza epidemics. Sick leave records may facilitate more timely detection of influenza outbreaks, as trends in increased sick leave may precede alerts issued by sentinel surveillance systems by days or weeks. Sick leave data have not been comprehensively evaluated in comparison to traditional surveillance methods. The aim of this paper is to study the performance and the feasibility of using a detection system based on sick leave data to detect influenza outbreaks.

Methods: Sick leave records were extracted from private French health insurance data, covering on average 209, 932 companies per year across a wide range of sizes and sectors. We used linear regression to estimate the weekly number of new sick leave spells between 2016 and 2017 in 12 French regions, adjusting for trend, seasonality and worker leaves on historical data from 2010 to 2015. Outbreaks were detected using a 95%-prediction interval. This method was compared to results from the French Sentinelles network, a gold-standard primary care surveillance system currently in place.

Results: Using sick leave data, we detected 92% of reported influenza outbreaks between 2016 and 2017, on average 5.88 weeks prior to outbreak peaks. Compared to the existing Sentinelles model, our method had high sensitivity (89%) and positive predictive value (86%), and detected outbreaks on average 2.5 weeks earlier.

Conclusion: Sick leave surveillance could be a sensitive, specific and timely tool for detection of influenza outbreaks.

Keywords: Surveillance, Influenza, Sick-leave, Outbreak detection

Background

Early outbreak detection is crucial for preparedness and timely public health and medical responses. It provides useful information to physicians, companies, and the public, ensuring proper drug prescription, health service planning, workplace preparedness, and continuity of operations in case of high absenteeism [1], among many other uses.

Most countries face periodic influenza (or "flu") epidemics that vary in size and severity from year to year [2]. Seasonal flu can be highly virulent and, like many respiratory viruses, can spread rapidly through populations highlighting a need for a robust epidemiological surveillance system to detect emerging outbreaks. Surveillance system guidelines developed by the US Centers for Disease Control and Prevention (CDC) suggest that systems should be simple, reliable, flexible, timely, and readily accepted by diverse individuals and organizations to ensure participation [3].

* Correspondence: tom.duchemin@cnam.fr
[1]MESuRS laboratory, Conservatoire National des Arts et Métiers, 292 Rue Saint-Martin, 75003 Paris, France
[2]Malakoff Humanis, 21 Rue Laffitte, 75009 Paris, France
Full list of author information is available at the end of the article

© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Flu surveillance systems vary by country and rely on various types of data. National health agencies monitor flu epidemics using healthcare records, medical sentinel systems, pharmaceutical sales and other data sources. Most of these systems rely on data from healthcare settings that rely on patient healthcare seeking behavior or after results of clinical tests, which often reflects those with symptoms or relatively advanced stages of the disease. These systems fail to capture individuals who do not seek medical care, whether due to asymptomatic infection, perceived mildness of infection or a general reluctance to seek care [1, 4]. Given these gaps, alternative data streams from non-healthcare settings may provide a valuable complement to classical surveillance systems [5].

Human resources data collected at the workplace have received relatively little attention for outbreak detection, but present characteristics that are useful for infectious disease surveillance. In many settings, absenteeism data are routinely collected and centralized either for the use by companies themselves or for health insurance purposes. Due to legal purposes and to the implication of work absence on salaries, these data are comprehensive and reliable. Though a handful of studies have assessed the role of sick leave data for outbreak detection, they did not develop a comprehensive assessment of its performance and robustness. Bollaerts et al., Patterson et al. and Groenewold et al. [1, 6, 7] assessed the usefulness of work absenteeism surveillance as a tool for early warning systems for influenza. Work absenteeism study can also supplement more traditional medical data by providing information about an epidemic's socioeconomic impact [1, 4]. As a consequence, the US National Institute for Occupational Safety and Health (NIOSH) has been monitoring health-related workplace absenteeism among full-time workers using data received monthly from the Current Population Survey since 2017 and making this data available online [8].

A great challenge in epidemiological surveillance lies in identifying data streams that allow for sensitive, specific and timely outbreak detection. In the context of outbreak detection, surveillance sensitivity can refer to both (i) the proportion of true cases detected, and (ii) the probability of detecting an outbreak, including the changes in the number of cases over time [3]. Surveillance specificity refers to the probability of correctly identifying when an outbreak is not occurring [9]. Lastly, surveillance timeliness generally refers to the time difference between an event and its standard reference [5]. Some studies have suggested that absenteeism data along with others such as over-the-counter pharmaceutical sales and emergency visits seem to be more timely than sentinel Influenza-Like Illness (ILI) surveillance [5, 6], other traditional flu data sources [7], physician diagnoses [5], and virological data [5].

In this study, we assess the sensitivity, positive predictive value and timeliness of workplace absenteeism data for detection of flu outbreaks in France. Our hypothesis is that monitoring sick-leave data at the workplace might help anticipate outbreaks in a timely manner using routinely collected data. We then compare the performance of a sick leave based monitoring system to the performance of the national standard surveillance system of influenza in France which is based on ILI data.

Methods

Data

Sick-leave data

The study relies on the sick leave record system of the French health insurance company Malakoff Médéric. Malakoff Médéric insures sick leaves for 114,707 to 245,973 French companies in a wide range of sectors, covering between 290,056 and 2,765,400 employees per year. This wide variation is due to the fact that some companies were no longer required to report these data to the insurer after 2015. The insured companies have on average 18.7 employees per year. Nearly half of the companies (46%) were in services, 36% in commerce, 12% in industry and construction, and 5% in health.

These data are routinely collected and annually reported in the system DADS (*Déclarations Annuelles de Données Sociales*). These data are reported by companies for payroll administration and is then transferred to the appropriate agencies (such as insurers) for administrative purposes. For our purposes, we used the weekly incidence rate of sick leave spells (per 100,000 workers) aggregated at the regional level across the 12 administrative regions of metropolitan France over the period 2010–2017.

Workers leave data

The number of workers on non-sick leave (e.g. paid holiday) is not reported in DADS, so the denominator of the weekly sick leave incidence rate was defined as the number of workers actively employed by their company during the observed week, and not the number of workers actually working during the week. To adjust our data, we used data from statistics department of the French Ministry of labour (DARES) to build an indicator (the worker-leave-peak indicator) describing weeks with a peak in non-sick leave [10]. Peaks were identified during the Christmas school holidays (last week of December and first week of January) and during summer (second week of July to third week of August).

Influenza-like illness data

Weekly sick leave incidence was compared to weekly ILI incidence (per 100,000 inhabitants), derived from the French influenza surveillance database of the GP Sentinelles network, coordinated by Santé Publique France. In 2018, the Sentinelles network was composed of 1314 general private practitioners and 116 private pediatricians, all voluntary participants and spread widely across the whole of France's territories. Detailed information on this network can be found elsewhere [11]. This ILI incidence data is the main source of data used to declare influenza epidemics in France. ILI are defined by Sentinelles as a fever above 39 °C, with sudden onset, accompanied by myalgia and respiratory signs. We can visually identify the ILI peak each year as the week where the ILI rate is highest between the 26th week (middle of the year) of year t and the 26th week of year t+ 1.

In addition to providing weekly ILI incidence data, the French Sentinelles network also proposes an ILI-outbreak detection algorithm. The algorithm is based on a Serfling method [12]. It has been adapted for routine surveillance of epidemics of ILI in France [13]. The method implemented by Sentinelles is based on a periodic regression model including a biannual seasonal effect of clinical influenza, a linear trend and an intercept adjusting for a baseline diagnostic activity (which corresponds to the number of influenza syndromes that would be diagnosed in the absence of influenza virus during the off-season).

This study exclusively relied on non-identifying, aggregated data for which no specific authorisation or ethical clearance was required. The sick-leave data are owned by Malakoff Humanis and we obtained authorization for the analyses.

Methods

Identification of influenza outbreak episodes

Dates of influenza outbreak episodes from Sentinelles are publicly unavailable so we trained the model described above on data from 1984 to 2009 to mimic the sentinel system. Weekly ILI incidence rates per 100,000 residents and per region from 2010 to 2017 were then compared to an outbreak detection threshold. This was defined as the upper bound of the 95% prediction interval from this model, and to increase specificity an alert was only declared when this threshold was crossed twice consecutively.

As reported in the literature [12, 14], we find here that the model correctly identifies all ILI peaks and does not detect false outbreak.

Determination of sick leave outbreak episodes

To detect sick leave outbreak episodes, an algorithm based on the Serfling method was used. The regression includes an intercept to adjust for the baseline sick-leave activity and the worker-leave-peak indicator to adjust for seasonality.

Similarly, to the Sentinelles method, an outbreak was declared if the true sick leave incidence rate crossed the 95% prediction interval twice consecutively. An alert was lifted when the incidence fell below the threshold, again for two consecutive weeks [15].

To mimic a prospective study, the model was fitted on sick leave data from 2010 to 2015 and years 2016 and 2017 were used to evaluate model performance: only historical data were then used for outbreak detection. As timing of ILI outbreaks may vary geographically, analyses were conducted separately for each of mainland France's 12 administrative regions. For simplicity, some results were plotted in the main text for three regions only, chosen to reflect a North-South gradient (respectively Haut-de-France, Ile-de-France and Provence-Alpes-Côte d'Azur). Full results are included as supplementary results.

Criteria for assessing the proposed surveillance system

Evaluation criteria were selected to answer two questions: (i) Does the sick leave model efficiently detect ILI outbreaks? and (ii) How does it compare to the Sentinelles model? Results are presented for each French administrative region and are also aggregated at the national level.

Performance of the sick leave model to detect ILI outbreaks To answer the first question, we calculated sensitivity and the positive predictive value to evaluate whether our model correctly detected all ILI outbreaks. The two criteria are:

$$Sensitivity_{per\ episode} = \frac{Number\ of\ Influenza\ outbreak\ peaks\ detected\ by\ the\ Sick\ Leave\ model}{Number\ of\ Influenza\ outbreaks}$$

$$Positive\ Predictive\ Value = \frac{Number\ of\ Sick\ Leave\ outbreak\ peaks\ crossing\ an\ Influenza\ outbreak}{Number\ of\ Sick\ Leave\ outbreaks}$$

We calculated a positive predictive value rather than a specificity since a specificity would require weeks with and weeks without outbreaks. This is impossible since outbreaks are only defined by a model, Sentinelles in our case, and this model is imperfect. We therefore do not have no outbreak episodes but only the peaks of those outbreaks that are identifiable without a model.

To evaluate our model's timeliness, we calculated the detection time we defined as the delay between the outbreak detection of our algorithm compared to the annual influenza outbreak peak. The influenza outbreak peak was defined as the week with the highest number of reported ILI cases between June 1st and May 31st of the subsequent year.

Performance of the sick leave model compared to the Sentinelles model To evaluate the performance of the sick leave model, we calculated its sensitivity and specificity with respect to the Sentinelles model. Unlike the previous criteria, we define these indicators at the level of the week rather than the episode. The objective is to assess whether our two models are similar. The two criteria are defined as follows:

$$Sensitiviy_{Sentinelles} = \frac{Number\ of\ weeks\ with\ a\ Sentinelles\ alert\ and\ with\ a\ Sick\ Leave\ alert}{Number\ of\ weeks\ with\ a\ Sentinelles\ alert}$$

$$Specificity_{Sentinelles} = \frac{Number\ of\ weeks\ with\ no\ Sentinelles\ alert\ and\ with\ no\ Sick\ Leave\ alert}{Number\ of\ weeks\ with\ no\ Sentinelles\ alert}$$

We also computed a Youden-index, an indicator that combines both sensitivity and specificity:

$$Younden\ index = Sensitivity_{Sentinelles} + Specificity_{Sentinelles} - 1$$

The timeliness of the sick-leave model compared to the Sentinelles model was also evaluated: the delay between the outbreak detection of the first model compared to the second one is computed.

For the sake of clarity and because the Sentinelles model is also imperfect, the comparison to the gold-standard Sentinelles will be presented in the Supplementary Material.

Results

Incidence curves obtained from the Sentinelles surveillance networks from 2010 to 2017 reveal annual peaks of influenza-like illness (ILI), from 163 to 1290 per 100,000 per week, occurring approximately between December and February (Figs. 1 and S1). During the summers, the incidence approaches zero. By comparison, weekly sick leave incidence varied about an annual average of 1021 to 1335 per 100,000 per week, depending on the region (Figs. 1 and S1). They exhibit greater variability and more peaks per year than ILI incidence. However, based on visual inspection, the highest seasonal peaks tend to coincide with ILI incidence peaks. Moreover, most of the seasonal sick-leave troughs coincide with Christmas and summer school holidays periods, which can be explained by a decrease of the at-risk population, i.e. an increase in the number of workers on paid leave. Finally, there is no apparent change between 2014 and 2015 despite the strong variation in the volume of workers in the database.

For three French regions, Fig. 2 presents the incidence of sick leaves and ILI for the 2015–2017 time period. In each region, the Sentinelles surveillance system identified exactly one alert per year, triggered a few weeks before or during the peak of ILI incidence (Figs. 2 and S2). The exception was Bretagne, where no alert was identified during winter 2016–2017 (Figure S2). By comparison, the sick leave surveillance system triggered one to three alert episodes per year.

We assessed the sick leave surveillance system on its ability to detect and anticipate ILI incidence peaks. Table 1 summarizes the indicators of its performance regarding ILI peak detection and anticipation in each region, averaged over the 2 years of the model test. The sensitivity per episode (probability of detection of the ILI outbreak) had a mean of 0.92 (range 0.5–1) across regions, while the positive predictive value per episode had an average of 0.58 (range 0.2–1). The sick leave alert generally occurred prior to the peak ILI incidence, on average 5.88 weeks (range 2.5–11) before the peak.

We also compared the performance of our sick leave model with the Sentinelles surveillance system. The results can be read in Table S1 (Supplementary Material).

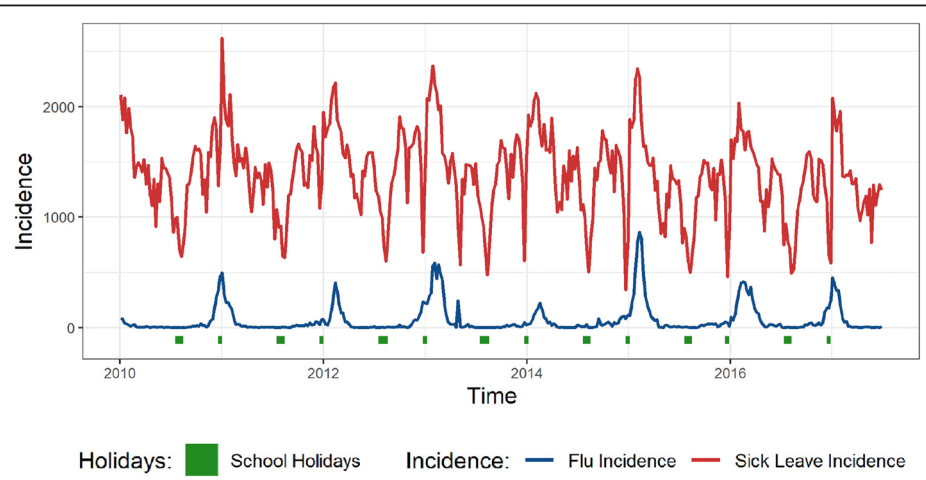

Fig. 1 Incidence of influenza-like illness and sick leave in Ile-de-France. Incidence per 100,000 per week of influenza-like illness and sick leave in Ile-de-France, the most populous region in France, 2010–2017. Christmas and summer school holidays (increased worker leave periods) are shown at the bottom

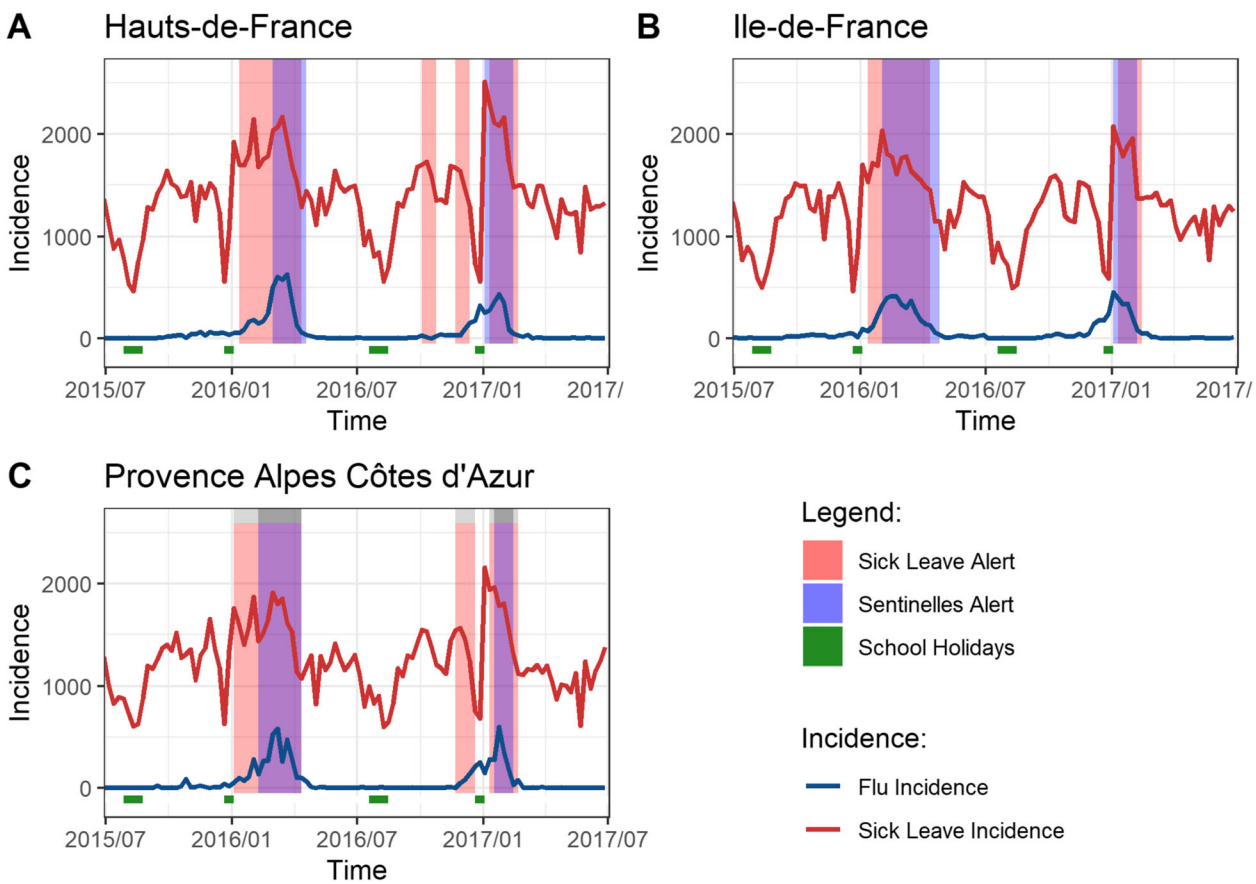

Fig. 2 Incidence of influenza-like illness and sick leave and results from Sentinelles and Sick-Leave models. Incidence per 100,000 per week of influenza-like illness and sick leave, 2015–2017, and alerts from the Sentinelles (blue) and the sick-leave (red) models, in three French regions (A: Hauts-de-France, B: Ile-de-France, C: Provence-Alpes-Côte d'Azur). Alerts show as purple when Sentinelles and sick-leave alerts overlap. Christmas and summer school holidays (increased worker leave periods) are shown at the bottom (green)

Table 1 Performance of the sick-leave model to detect ILI outbreaks

	Sensitivity per episode	Positive predictive value	Detection time (weeks) before ILI peak
Auvergne-Rhone-Alpes	0.5	0.2	7
Bourgogne-Franche-Comte	0.5	0.2	11
Bretagne	1	1	2.5
Centre-Val de Loire	1	0.5	6.5
Grand Est	1	0.5	5
Hauts-de-France	1	0.5	7
Ile-de-France	1	1	3
Normandie	1	0.33	6
Nouvelle-Aquitaine	1	0.5	6.5
Occitanie	1	1	3
Pays de la Loire	1	0.5	6.5
Provence-Alpes-Cote-d'Azur	1	0.66	6.5
Total mean	0.92	0.58	5.88

For each region, the values of these indicators are averaged over the 2 years evaluated

The main result is that the sick leave model alert was always triggered earlier than the Sentinelles model when both alerts match, with an average lead time of 2.5 weeks (range: 0.5–4).

Discussion

Workplace absenteeism data can be used by public health surveillance systems to detect emerging infectious disease epidemics [16]. Despite this, to this date, few health authorities worldwide use absenteeism data to inform outbreak surveillance. Here, we assessed the potential of workplace absenteeism data to monitor influenza and detect epidemics in France, using an adapted statistical method to analyze this data. We applied this method to a comprehensive national database of workplace absenteeism and validated it against the French national surveillance system based on sentinel GPs. Our results suggest that a system based on workplace absenteeism could be highly sensitive and detect influenza epidemics earlier than the current French surveillance system.

We found that the surveillance system we propose would be able to detect outbreaks 5.9 weeks before the peak and about 2.5 weeks before the Sentinelles system. This suggests that sick-leave data could be almost as timely as emergency visits data that has a timeliness of 3 weeks compared to ILI data systems [17]. Our findings in this regard are in line with previously published studies in several contexts worldwide. In a French study from 1994, sick-leave data from a large company allowed to detect flu epidemics with up to 2 weeks of advance [9]. In a more recent Belgian study, worker absenteeism data from the Belgian Medical Expertise and from the Belgian railway system was shown to start rising 2–3 weeks in advance and to peak 2 weeks in advance, as compared with ILI data from the Belgian sentinel GP surveillance system [6]. In the UK, data on workplace absenteeism among employees of Transport for London peaked up to 2 weeks before the NHS ILI surveillance data [7]; and monitoring workplace absence due to “cold”, “cough” or “influenza” among the staff of a large hospital organization was shown to allow the detection of flu epidemics with a significant advance of up to 9 weeks [18].

Very few of these previously published studies included an assessment of the sensitivity, specificity or positive predictive value of an absenteeism-based surveillance system. However, in the French study from 1994, the sensitivity and specificity of surveillance based on sick-leave data from a large company were estimated at 74 and 67% for the identification of epidemic weeks (Youden index: 0.42) and 67 and 94% for the detection of epidemics (Youden index: 0.61), with an 80% positive predictive value [9]. The UK study based on hospital staff absenteeism also noted that the resulting system did not lead to more false positives than the NHS surveillance data in London.

The quality of the developed model depends strongly on the quality of the data collected within companies. Sick-leave data have the advantage of describing quasi-real individual behavior regarding sick-leave and presence at the workplace. These data are in fact used to enter employees’ pay and are subsequently fulfilled by obligation in the computer system. Our data may not be representative of French population because it describes data from a health-insurer. For instance, it insures few construction companies because they have their own specific insurer. The data also do not include unemployed people. However, representativeness is not necessarily required to build up an outbreak detection system. In fact, outbreak detection aims to detect any unusual, expected number of cases to generate signal alarms. Similarly, the GP Sentinelles network does not include all GPs but a small subset of the same practitioners over time.

Another downside of our data is related to the definition of the sick leave rate. The denominator of this rate is the number of workers and it includes workers that are on holiday. The sick leave rate therefore drops during school holidays and the systems do not detect any alerts during those holidays even if they are included in the statistical model as covariates. This is an issue if the epidemic occurs during the holidays and this is the case during the year 2016–2018 where the peak occurs during the Christmas break for two regions. The model could then be more sensitive if the denominator was the number of employees actually at work.

Furthermore, the model accuracy estimated by the false positive rate is strongly related to the definition of cases to detect. The time series of cases is based only on flu cases. However, other epidemics such as gastroenteritis can influence sick leaves data and could be considered for future works: the sick leaves outbreaks may correspond to other disease and may explain the poor positive predictive value in some cases. The detection algorithm could also be improved if the estimation of expected cases could adjust for any potential past exogenous environmental factors such as a terrorist attack, strikes or unexpected bad weather episodes. The model accuracy is moreover strongly related to the method chosen for the algorithm. The Serfling method may actually not be the more accurate model and was chosen to be consistent with the Sentinelles method. Some other regression-based models are known to be more specific, like the Farrington algorithm [19, 20].

Another limitation of our study is linked to the fact that we relied on data that were consolidated on an annual basis, and not in real-time. As such, the system is

able to detect outbreak based on the dates of sick-leaves, but only retrospectively. A proper integration of sick-leave data into a health surveillance system would thus require an effort to ensure sick-leave data are consolidated and made available in real time. Our results suggest that the resulting increased timeliness of a surveillance system including this data stream may justify this effort. Moreover, these data sometimes already exist in near-real time because of local legislation. For instance, French companies must declare sick leave within 5 days.

Sick leave could then be an alternative source of data for the detection of sick leave. Many other data sources have already been used for near real-time prediction of influenza outbreaks: GP networks of course, but also Google queries with Google flu [21], social network database [22], large epidemiological databases [23] and, more recently, wastewater have been used for COVID-19 monitoring [24]. Sick leave data is however more robust than Google (and probably social network) data, which has shown its limitations. Google queries for relevant keywords are indeed not only correlated to the epidemic but is also correlated with Google's practices: the data generation process changes over time and Google sometimes recommends certain keywords for commercial purposes [25]. Moreover, large epidemiological database may be more sensitive and specific but are much more complex to handle. Sick leave data can be simple to retrieve, easy to analyze and most of all very timely compared to the data usually used. These data could thus enable near real-time prediction, which would allow for reactive monitoring of influenza outbreak.

Conclusion

Many of the previously published studies assessing the potential of workplace absenteeism data for flu surveillance simply provided visual analyses comparing and correlating absenteeism data with ILI surveillance data [6, 7]. By contrast, in this work, we propose a statistical approach and algorithm to analyze the French workplace absenteeism data, and to raise alarms when outbreaks are detected. This allows us to both propose a complete surveillance system that could be used in practice provided the data is available, and to fully assess the performance of this surveillance system.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-020-05754-5.

Additional file 1. Figure S1: Incidence per 100,000 per week of influenza-like illness and sick leave in twelve French regions, 2010–2017. The Christmas and summer school holidays (increased worker leave periods) are shown at the bottom. **Figure S2:** Incidence of influenza-like illness and sick leave, 2015–2017, and alerts from the Sentinelles and the sick-leave models, in twelve French regions. The Christmas and summer school holidays (increased worker leave periods) are shown at the bottom. **Table S1:** Performance of the sick-leave model compared to the Sentinelles model.

Abbreviations

ILI: Influenza-like illness; GP: General practitioner

Acknowledgements

Not applicable.

Authors' contributions

TD, JB, RL: analyzed and extracted the data. TD, JB, PAAT, WD, KJ, LT, MH: Wrote the first draft of the paper. TD, JB, PAAT, RA, OSD, AD, JPG, NN, HN, DRMS, WD, KJ, LT, MNH: designed and discussed statistical analysis and interpreted data. All authors: Read, reviewed and approved the final manuscript.

Funding

TD PhD is funded by Association Nationale de la Recherche et de la Technologie and Malakoff Humanis.
JB PhD is funded by the INCEPTION project (PIA/ANR-16-CONV-0005).
PAAT PhD is funded by INSERM-ANRS (France Recherche Nord & Sud Sida-HIV Hépatites), grant number ANRS-12377 B104.
DS PhD is funded by a Canadian Institutes of Health Research Doctoral Foreign Study Award (Funding Reference Number 164263) as well as the French government through its National Research Agency project SPHINX-17-CE36–0008-01.

Availability of data and materials

The data that support the findings of this study are available from Malakoff Humanis but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Malakoff Humanis.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

TD and RL are employed of Malakoff Humanis.

Author details

[1]MESuRS laboratory, Conservatoire National des Arts et Métiers, 292 Rue Saint-Martin, 75003 Paris, France. [2]Malakoff Humanis, 21 Rue Laffitte, 75009 Paris, France. [3]Institut Pasteur, Epidemiology and Modelling of Antibiotic Evasion (EMAE), Paris, France. [4]PACRI unit, Conservatoire National des Arts et Métiers, Institut Pasteur, Paris, France. [5]Université Paris-Saclay, UVSQ, Inserm, CESP, Anti-infective evasion and pharmacoepidemiology team, Montigny-Le-Bretonneux, France. [6]Biodiversity and Epidemiology of Bacterial Pathogens, Institut Pasteur, Paris, France. [7]Malaria: Parasites and Hosts, Department of Parasites and Insect Vectors, Institut Pasteur, Paris, France.

Received: 8 July 2020 Accepted: 28 December 2020
Published online: 11 January 2021

References

1. Groenewold MR, Konicki DL, Luckhaupt SE, Gomaa A, Koonin LM. Exploring National Surveillance for Health-Related Workplace Absenteeism: Lessons Learned From the 2009 Influenza A Pandemic. Disaster Med Public Health Preparedness. 2013;7(2):160–6.
2. World Health Organization. WHO Fact sheets, Influenza (Seasonal). 2018 [cited 2020 May 18]. Available from: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal).
3. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, et al. Updated guidelines for evaluating public health surveillance systems:

recommendations from the Guidelines Working Group. MMWR Recomm Rep. 2001;50(RR-13):1–35 quiz CE1–7.
4. Groenewold M, Burrer S, Ahmed F, Uzicanin A. National Surveillance for Health-Related Workplace Absenteeism, United States 2017–18. Online J Public Health Inform. 2019;11(1) [cited 2020 May 18]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6606163/.
5. Dailey L, Watkins RE, Plant AJ. Timeliness of Data Sources Used for Influenza Surveillance. J Am Med Inform Assoc. 2007;14(5):626–31.
6. Bollaerts K, Antoine J, Robesyn E, Van Proeyen L, Vomberg J, Feys E, et al. Timeliness of syndromic influenza surveillance through work and school absenteeism. Arch Public Health. 2010;68(3):115–20.
7. Paterson B, Caddis R, Durrheim D. Use of Workplace Absenteeism Surveillance Data for Outbreak Detection. Emerg Infect Dis. 2011;17(10):1963–4.
8. NIOSH - CDC. Absenteeism in the Workplace. 2020 [cited 2020 May 18]. Available from: https://www.cdc.gov/niosh/topics/absences/default.html.
9. Quenel P, Dab W, Hannoun C, Cohen JM. Sensitivity, Specificity and predictive Values of Health Service Based Indicators for the Surveillance of Influenza A Epidemics. Int J Epidemiol. 1994;23(4):849–55.
10. DARES. Les congés payés et jours de RTT : quel lien avec l'organisation du travail ? DARES Analyse. 2017; [cited 2020 May 18]. Available from: https://dares.travail-emploi.gouv.fr/IMG/pdf/2017-054.pdf.
11. Valleron AJ, Bouvet E, Garnerin P, Ménarès J, Heard I, Letrait S, et al. A computer network for the surveillance of communicable diseases: the French experiment. Am J Public Health. 1986;76(11):1289–92.
12. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep. 1963;78(6):494–506.
13. Costagliola D, Flahault A, Galinec D, Garnerin P, Menares J, Valleron AJ. A routine tool for detection and assessment of epidemics of influenza-like syndromes in France. Am J Public Health. 1991;81(1):97–9.
14. Souty C, Jreich R, Le Strat Y, Pelat C, Boëlle PY, Guerrisi, et al. Performances of statistical methods for the detection of seasonal influenza epidemics using a consensus-based gold standard. Epidemiol Infect. 2018;146(2):168–76.
15. Retel O, Fortin N, Henry V, Hubert B, Faisant M, Casamatta D, et al. Contribution des associations SOS Médecins à une surveillance locale de la grippe saisonnière en France. Bull épidémiol hebd. 2014;28:466–72.
16. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V, CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR Recomm Rep. 2004;53(RR-5):1–11.
17. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerging Infect Dis. 2004;10(5):858–64.
18. Drumright LN, SDW F, Elliot AJ, Catchpole M, Pebody RG, Atkins M, et al. Assessing the use of hospital staff influenza-like absence (ILA) for enhancing hospital preparedness and national surveillance. BMC Infect Dis. 2015;15:110.
19. Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. Stat Med. 2013;32(7):1206–22.
20. Farrington CP, Andrews NJ, Beale AD, Catchpole MA. A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. J R Stat Soc Series A. 1996;159(3):547–63.
21. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza Forecasting with Google Flu Trends. PLOS ONE. 2013;8(2):e56176.
22. Alessa A, Faezipour M. A review of influenza detection and prediction through social networking sites. Theor Biol Med Model. 2018;15(1):2.
23. Santillana M, Nguyen AT, Louie T, Zink A, Gray J, Sung I, et al. Cloud-based Electronic Health Records for Real-time, Region-specific Influenza Surveillance. Sci Rep. 2016;6(1):25732.
24. Mao K, Zhang H, Yang Z. Can a Paper-Based Device Trace COVID-19 Sources with Wastewater-Based Epidemiology? Environ Sci Technol. 2020;54(7):3733–5.
25. Lazer D, Kennedy R, King G, Vespignani A. The Parable of Google Flu: Traps in Big Data Analysis. Science. 2014;343(6176):1203–5.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC