BMC
Medical Research Methodology

RESEARCH ARTICLE   **Open Access**

A statistical model to assess the risk of communicable diseases associated with multiple exposures in healthcare settings

Cécile Payet[1,2*], Nicolas Voirin[1,2], Philippe Vanhems[1,2] and René Ecochard[2,3]

Abstract

Background: The occurrence of communicable diseases (CD) depends on exposure to contagious persons. The effects of exposure to CD are delayed in time and contagious persons remain contagious for several days during which their contagiousness varies. Moreover when multiple exposures occur, it is difficult to know which exposure is associated with the CD.

Methods: A statistical model at the individual level is presented to estimate the risk of CD to patients, in healthcare settings, with multiple observed exposures to other patients and healthcare workers and unobserved exposures to unobserved or unobservable sources. The model explores the delayed effect of observed exposure, of source contagiousness and of unobserved exposure. It was applied to data on influenza-like illness (ILI) among patients in a university hospital during 3 influenza seasons: from 2004 to 2007. Over a total of 138,411 patients-days of follow-up, 64 incident ILI cases were observed among 21,519 patients at risk of ILI.

Results: The ILI risk per 10,000 patients-days associated with observed exposure was about 129.1 (95% Credible Interval (CrI): 84.5, 182.9) and was associated at 72% with exposures to patients or healthcare workers 1 day earlier and at 41% with the 1st day of source contagiousness. The ILI risk associated with unobserved exposure was 0.8 (95% CrI: 0.3, 1.6) per 10,000 patients-days in non-epidemic situation in the community and 4.3 (95% CrI: 0.4, 11.0) in epidemic situation.

Conclusions: The model could be an interesting epidemiological tool to further assess the relative contributions of observed and unobserved exposures to CD risk in healthcare settings.

Keywords: Disease transmission, Infectious, Health facilities, Influenza, Human, Models, Statistical, Risk

Background

The occurrence of communicable diseases (CD) depends on exposure to contagious persons. This concept, called *dependent happenings* [1,2], should be considered in statistical models to assess CD risk, for example by incorporating explicative variables describing exposure [2].

Exposure to infection refers to the potentially infectious contacts a person has and some knowledge of natural disease history is needed to define exposure.

Once a person exhibits symptoms of CD, only exposures that have occurred in period compatible with the incubation period of the susceptible host may lead to transmission. Also, exposure to an infectious individual could only lead to a subsequent infection (or disease) if exposure happens during the infectious period of the infected person. But, infectious periods and incubation periods differ individually and neither the durations of these periods nor the level of viral shedding are generally known. However, timing of exposure and contagiousness of the source are two important determinants of the risk of infection that can be included in statistical models aiming to analyze infectious diseases data.

Considering the moment of exposure, the effects of exposures to CD are often delayed in time and CD

* Correspondence: cecile.payet@univ-lyon1.fr
[1]Hospices Civils de Lyon, Service d'Hygiène, Épidémiologie et Prévention, Unité Epidémiologie et Biomarqueurs de l'Infection, Lyon F-69437, France
[2]CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Epidémiologie et Santé Publique, Université de Lyon; Université Lyon 1, Villeurbanne F-69100, France
Full list of author information is available at the end of the article

occurring at a given time can potentially be attributed to past exposures at different times (i.e. lags). The effects of exposure are distributed over a specific time period and can be modeled with several parameters to explain the contribution of exposures at different lags.

Similarly, regarding the source, contagious persons remain contagious for several days during which their contagiousness varies. It is, therefore, expected that the effects of exposure also vary with source contagiousness, the effects being distribution over the days of the contagious period.

Consequently, the effects of exposure on CD risk can be modeled as the joint result of delayed effects associated with lags and effects associated with the source person's contagiousness, and a probabilistic approach may allow estimating the distributions of the risk according to lags and the source person's contagiousness.

We propose a statistical model at the individual level to assess CD risk, in healthcare settings, among patients who may be subjected to multiple exposures. In this context, all sources of infection (patients, HCWs, or visitors) are usually imperfectly identified. So, it is important to consider not only observed exposure to other patients and healthcare workers (HCWs) but also exposures to unobserved or unobservable sources such as symptomatic infectious persons unnoticed during observation, asymptomatic persons or external sources. Incidence of CD in the community can be used to represent the unobserved exposures. The proposed model explores the delayed effect of multiple observed exposures, of source contagiousness and of unobserved or unobservable exposures. It was applied to data on influenza-like illness (ILI) among patients in a university hospital.

Methods
Model
A patient can contract a CD during his/her stay in a healthcare setting. This hospital-acquired CD may result from exposure to contagious HCWs or contagious patients as well as contagious persons coming from outside the healthcare setting, such as visitors. In addition, these exposures are usually imperfectly identified in epidemiologic studies because generally clinical cases are of interest. Exposures can be observed (i.e. documented), unobserved (i.e. unnoticed) or unobservable (e.g. asymptomatic contagious persons). In the following, only the term "unobserved exposure" will be used but it must be remembered that it referred to "unobserved and unobservable exposure". Separating these two effects is out of the scope of the present work.

When a hospitalized person has a CD, multiple exposures delayed in time may be consistent with disease onset and it is difficult to know with certainty

which exposure is associated with the CD. The model is based on the concept that escaping to infection from each exposure is required to avoid infection. If $\lambda_{ij}^{observed}$ is the CD risk associated with observed exposure (w_{ij}), then the cumulative risk of escaping CD following a series of exposures is $\prod_{ij}\left(1 - \lambda_{ij}^{observed}\right)^{w_{ij}}$. In other words, to avoid CD, a patient should escape infection from all exposures.

To estimate the CD risk to patients, taking into account the effect of observed exposure associated with lags and the day of source person contagiousness as well as unobserved exposure, we constructed the model at the individual level, so that, for a given patient and day, CD risk λ can be modeled as follows:

$$y \sim Bernoulli \; (\lambda)$$
$$\lambda = 1 - \left(1 - \lambda^{unobserved}\right)\prod_{ij}\left(1 - \lambda_{ij}^{observed}\right)^{w_{ij}}$$

where $y = 1$ if the patient has a CD and $y = 0$ if not, w_{ij} is observed exposure, equals 1 if the patient has been exposed to patients or HCWs i days earlier at their j^{th} contagiousness day and $w_{ij} = 0$ if not, $\lambda_{ij}^{observed}$ is the CD risk associated with observed exposure, and $\lambda^{unobserved}$ is the CD risk associated with unobserved exposure. As represented by y, the model includes ILI cases and non-cases. Patients were followed since admission at hospital. Censoring occurred for cases at the time of ILI and non-cases were followed up to discharge. So each admitted patient contributed to the risk set, until onset of ILI for cases and until discharge for non-cases. In addition, ILI patients were assumed to be immunized after ILI and therefore no more at risk.

To define maximum lag, we assume a maximum incubation period of the CD of I days. Thus, exposure of a patient can be defined as the presence of a contagious person in the same ward during the preceding I days. Contagious persons remain contagious for a given period and in the following, we assume a maximum contagiousness period of J days.

Observed exposure
For capturing the effects of observed exposure, we used an approach belonging to the context of the transmission models based on "who acquires infection from whom" (WAIFW) matrices , for which estimation methods are available elsewhere [3,4]. The CD risk observed exposure can be decomposed as follows (see Additional file 1):

$$\lambda_{ij}^{observed} = 1 - \exp\left(-k\,a_i\,b_j\right)$$

with constraints $\sum_i a_i = 1$ and $\sum_j b_j = 1$, where k is a scale parameter representing the CD risk, $[a_1, \ldots, a_i, \ldots,$

$a_I]^{\mathrm{T}}$ being the distribution vector of this risk over the preceding i days, and $[b_1,...,b_j,...,b_J]$ being the distribution vector of risk over j days of source contagiousness. This decomposition in 2 dimensions allows easy interpretation of parameters and makes the assumption of independence between the effect of lags and of the day of contagiousness. It also supposes that the risk is constant a given day for a given individual which is plausible since exposure is not expected to vary in this interval.

Unobserved exposure
In the model, we allowed exposures to unobserved sources such as exposures to symptomatic contagious persons unnoticed during observation, asymptomatic persons or external sources. We further assumed that unobserved exposure depended on the community incidence of CD. The CD risk associated with unobserved exposure in a healthcare setting can then be written as follows:

$$\lambda^{unobserved} = \exp\left(\beta^{Non-epidemic} + \beta^{Epidemic}x\right)$$

where $x = 1$ if the epidemic threshold has been reached on a given day in the community and $x = 0$ if it has not been reached. When CD incidence in the community is below the epidemic threshold, $\lambda^{unobserved} = \exp(\beta^{Non-epidemic})$. This risk is increased by factor $\beta^{Epidemic}$ when the incidence in the community is above the epidemic threshold.

Inference methods
We combined two methods for inference. First, maximum likelihood estimates of the model parameters were estimated as well as their 95% confidence intervals (95% CI) obtained from the Hessian matrix. The optim function in R was used. Then, the number of cases being small, a Bayesian inference was performed to obtain more robust intervals estimates. MCMC (*Markov Chain Monte Carlo*) methods [5] gave posterior parameters means and 95% credible intervals (95% CrI) from 2.5% and 97.5% quantiles of posterior distribution. Parameters were assigned non-informative priors [6]. Three independent MCMC with 3 different initial values were run in parallel to assess convergence on posterior parameter distributions and to check that the results were not sensitive to the choice of starting values. Gelman and Rubin convergence criteria [7] was also calculated and examined. After an initial burn-in period of 20,000 iterations, 80,000 iterations served to compute posterior parameter distributions. MCMC were analyzed with OpenBugs and the coda package in R.

Exposure to ILI in healthcare settings
The model was applied to assess the effects of exposure on ILI risk in healthcare settings. The data originated from 3 sources. The first source was a prospective observational study carried out between November 15, 2004 and April 15, 2007 at Edouard Herriot Hospital in Lyon, France [8]. During the study, each participating ward was followed-up daily to detect ILI cases. A case was defined as an adult patient or HCW presenting with fever (≥37.8°C) and cough or sore throat. A non-case was defined as a patient or a HCW free from ILI during the study period. As secondary source, all hospital data (admission/discharge dates and ward) on cases and non-cases were extracted from the hospital's information system. Data on HCW work periods were included. To quantify exposure to ILI and to define lags, the maximum incubation period was set at $I = 5$ days [9]. Therefore, for each patient, exposure was defined as the presence of a contagious person in the same ward and on 5 previous days. We assumed a contagious period for ILI of $J = 6$ days, starting 1 day before symptom onset [9]. Having a small number of cases, exposures to contagious persons 4 and 5 days earlier were grouped together. Similarly, exposures occurring during the 4th to 6th days of the source contagiousness were pooled. The third source was the French Institute for Public Health Surveillance [10], for defining epidemic and non-epidemic periods in the Rhône-Alpes region.

We assumed that infected patients didn't become susceptible again and therefore follow-up was censored the day of ILI onset for cases and the day of discharge for patients free from ILI. Over a total of 138,411 patient-days (19,773 weeks) of follow-up, 64 incident ILI cases were observed among 21,519 patients at risk of ILI.

Results
Parameter estimation by maximum likelihood and Bayesian inference gave similar results (Table 1).

The daily ILI risk per 10,000 patients-days associated with observed exposure was 129.1 (95% CrI: 84.5, 182.9). The ILI risk associated with unobserved exposure was 0.8 (95% CrI: 0.3, 1.6) in non-epidemic situation in the community and 4.3 (95% CrI: 0.4, 11.0) in epidemic situation. The distribution of ILI risk according to lags was 72%, 6%, 12% and 10% for observed exposures occurring 1, 2, 3 and 4–5 days earlier respectively. The distribution of this risk was 41%, 19%, 25% and 14% during observed exposure to patients or HCWs 1 day before their symptom onset, 1st, 2nd and 3rd to 5th days after symptom onset respectively.

Figures 1 and 2 present the daily ILI risk according to the day of contagiousness, lags and epidemic threshold in the community (see Additional file 1 for details of calculations). In the figures, each bar represents the risk of ILI associated with one exposure at a given lag and a given day of contagiousness, in the absence of all other exposures. The results indicate that ILI risk was highest

Table 1 Parameter estimates of Influenza-Like Illness (ILI) risk in a university hospital by 2 inference methods

|	Maximum likelihood		Bayesian inference	
	Mean	95% confidence interval	Mean	95% credible interval
ILI risk associated with unobserved exposure (per 10,000 patients-days)				
No epidemic in the community	0.8	0.3-1.6	0.8	0.3-1.6
Epidemic in the community	4.3	0.4-11.0	4.3	0.4-11.0
ILI risk associated with observed exposure (per 10,000 patients-days)	129.1	84.5-182.9	129.1	84.5-182.9
Distribution of risk according to lags				
1 day	82	51-95	72	52-89
2 days	1	0-98	6	0-21
3 days	15	0-81	12	1-32
4 to 5 days	2	- *	10	0-27
Total	100	-	100	-
Distribution of risk according to the source's contagious period				
1 day before the onset of source symptoms	47	24-71	41	21-62
1st day after the onset of source symptoms	17	6-42	19	4-39
2nd day after the onset of source symptoms	24	8-51	25	9-44
3rd to 5th days after the onset of source symptoms	12	- *	14	4-29
Total	100	-	100	-

*Confidence interval not computable because this parameter was inferred from others.

when observed exposure took place the previous day whatever the day of contagiousness and ILI incidence in the community. A slight decrease in risk was observed with the day of contagiousness.

Discussion

In the present work, a statistical individual model allows estimation of CD risk to patients who may have had multiple observed exposures and unobserved exposures in healthcare settings. The model is defined at the patient and daily levels, which allowed capturing individual and temporal variations of observed and unobserved exposures to finally precisely assess the relationship between exposures and CD risk. An additional feature of the model is to integrate the natural history of the disease. Then, the model can serve to identify days of exposure associated with
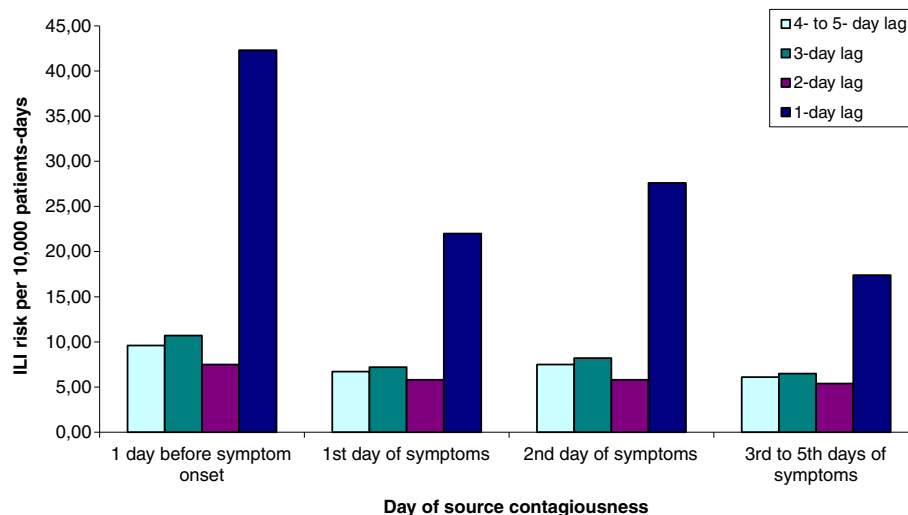
Figure 1 Risk of Influenza-Like Illness (ILI) associated with exposure to a patient or a healthcare worker according to a given lag and day of source contagiousness in a healthcare setting during community epidemic period. Each bar represents the risk of ILI associated with one exposure at a given lag and a given day of contagiousness, in the absence of all other exposures.
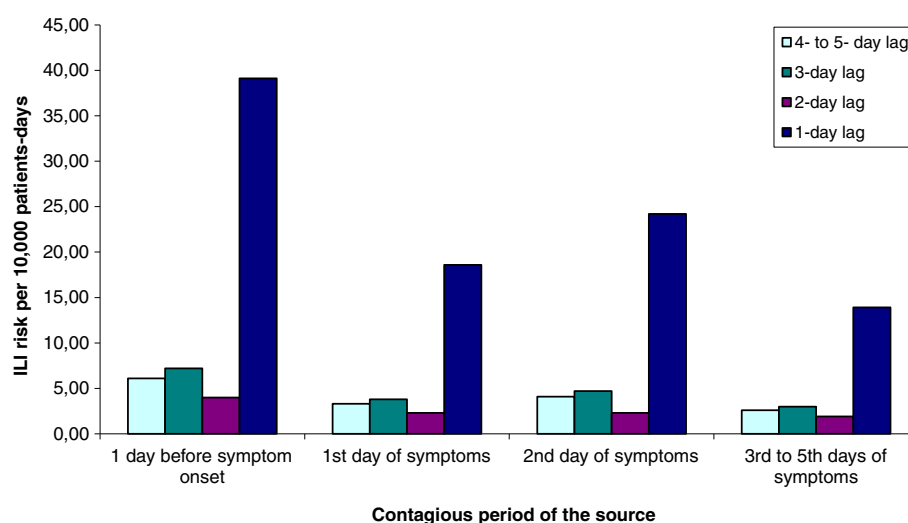
Figure 2 Risk of Influenza-Like Illness (ILI) associated with exposure to a patient or a healthcare worker according to a given lag and day of source contagiousness in a healthcare setting outside community epidemic period. Each bar represents the risk of ILI associated with one exposure at a given lag and a given day of contagiousness, in the absence of all other exposures.

increased risk of CD as well as days of contagiousness associated with heightened risk of disease transmission.

Application of this model to ILI showed that, on any given day, the observed ILI risk among adult patients was 129.1 per 10,000 patients-days and was mainly associated with exposures to patients or HCWs 1 day earlier and during the 1st day of source contagiousness. These results suggest 1) that the incubation period of ILI could be ≤24 h, and 2) that ILI could be transmitted when the source has no symptoms yet (for ILI, contagiousness starts 1 day before symptom onset [9]). In addition, the ILI risk appeared to decrease gradually, depending on the day of contagiousness. These results are consistent with current knowledge on ILI [9,11], which supports potential application of the model in hospital epidemiology.

These results indicate that prevention measures, such as closing wards, restricting visits, avoiding contact between healthy patients and contagious patients or HWCs, isolating contagious persons, or antiviral treatment and prophylaxis, should be started as soon as a case is suspected. A delay in alert or implementation of these measures could increase the risk of transmission. In addition, contact tracing may be useful to identify patients who may have been in contact with a contagious symptomatic patient or HCW the day preceding his or her symptom onset. These results are proposed for ILI and influenza among adult patients, but the model could be easily extended to children with longer contagious periods, to different subtypes of influenza viruses with various natural histories, and to other infections sharing similar routes of transmission and natural histories.

In comparison to observed exposure, unobserved exposure accounted for a small part of the total transmission. This may suggest that contagious persons unnoticed during observation, asymptomatic persons or external sources may play a limited role in ILI transmission in healthcare settings, compared to symptomatic sources. It is unlikely that symptomatic persons or external sources were missed because the ILI definition we used had a high sensitivity. A lower infectiousness of asymptomatic sources may also explain this slight effect. However, interestingly, the effect of unobserved exposure seems to increase, by a 5-fold factor, during epidemic period compared to non-epidemic period. In absence of an active surveillance, this result is in favor of limiting exposures during epidemic period or during emergencies such as a pandemic. It could be also an argument to promote influenza vaccination for patients as well as for HCWs.

The approach presented here had some limitations. Exposure was defined as the presence of a contagious person in the same ward and this definition ignored information on actual contacts between persons. In practice, however, it is difficult to follow each individual to find out with whom, when and for how long he/she came into contact with other individuals. In addition, the definition of exposure may seem more appropriate than contact because it includes all ILI transmission types. We did not follow patients outside hospital and it is possible that some patients presented ILI just after discharge. In this case, it is possible that transmission was underestimated. The bias was limited, the statistical model taking into account the censoring process, i.e. the discharge of the patients. The model was based on the assumption of independence between the lag of exposure and the day of source contagiousness. While this assumption appeared to be weak in our case, there were

other ways to estimate the parameters of a WAIFW matrix, for example, by equality constraints between parameters [3]. However, this hypothesis had the advantage of facilitating the estimation and interpretation of parameters. Equality constraints seem to be not entirely appropriate and could limit study of the influence of the day of contagiousness and of the lag of exposures on CD risk. However, shape constraints [12], reflecting the natural history of disease or adding *a priori* information, should be considered in future improvements of the model. For example, incubation [9,11] and contagiousness periods [9] follow upswing and downswing trends, and it would be interesting to consider such information in the model.

Due to the hierarchical structure of the hospital (i.e. patients in wards and wards in the hospital), it is likely that patients in the same wards may be not independent. A mixed effects model would be an interesting extension of the model, by adding a ward-specific scale parameter (k) representing a possibly different CD risk for each participating ward in the study. However the low number of events per ward (some wards experienced no ILI) would not have permitted to obtain reasonable estimates under the mixed-effects version of the model. It would be relevant to apply this model and its extensions to larger datasets.

Conclusions

Our results disclosed that the presence of a contagious patient or HCW dramatically increased the risk of ILI to other patients. Therefore, the proposed statistical model could be an interesting epidemiological tool to further assess the relative contribution of observed and unobserved exposures to CD risk in healthcare settings. Beyond the statistical approach, these results will be helpful to improve daily prevention of CDs in hospitals.

Additional file

Additional file 1: Decomposition of CD risk associated with observed exposure.

Author details
¹Hospices Civils de Lyon, Service d'Hygiène, Epidémiologie et Prévention, Unité Epidémiologie et Biomarqueurs de l'Infection, Lyon F-69437, France. ²CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Epidémiologie et Santé Publique, Université de Lyon; Université Lyon 1, Villeurbanne F-69100, France. ³Hospices Civils de Lyon, Service de Biostatistique, Lyon F-69003, France.

References

1. Ross R: **An application of the theory of probabilities to the study of a priori pathometry. Part I.** *Proc R Soc London Series A* 1916, **92**:204–230.
2. Rhodes PH, Halloran ME, Longini IM: **Counting process models for infectious disease data: distinguishing exposure to infection from susceptibility.** *JR Statist Soc B* 1996, **58**:751–762.
3. Anderson RM, May RM: *Infectious Diseases of Humans: Dynamics and Control.* Oxford: Oxford University Press; 1991.
4. Roberts MG, Heesterbeek JAP: *Mathematical Models in Epidemiology.* Mathematical Models: Encyclopedia of Life Support Systems; 2003.
5. Gilks WR, Richardson S, Spiegelhalter DJ: *Markov Chain Monte Carlo in Practice.* London: Chapman & Hall; 1996.
6. Gelman A, Carlin JB, Stern HS, et al: *Bayesian Data Analysis.* 2nd edition. Boca Raton, FL: Chapman & Hall/CRC Press; 2004:182–184.
7. Gelman A, Rubin DB: **Inference from iterative simulation using multiple sequences.** *Stat Sci* 1992, **7**:457–511.
8. Vanhems P, Voirin N, Roche S, et al: **Risk of influenza-like illness in an acute health care setting during community influenza epidemics in 2004–2005, 2005–2006, and 2006–2007: a prospective study.** *Arch Intern Med* 2011, **171**(2):151–157.
9. Carrat F, Vergu E, Feguson NM, et al: **Time lines of infection and disease in human influenza: a review of volunteer challenge studies.** *Am J Epidemiol* 2008, **167**:775–785.
10. Site de l'Institut de Veille Sanitaire. Bulletin épidémiologique grippe. http://www.invs.sante.fr. (Accessed April 10, 2011).
11. Lessler J, Reich NG, Brookmeyer R, et al: **Incubation periods of acute respiratory viral infections: a systematic review.** *Lancet Infect Dis* 2009, **9**:291–300.
12. Silvapulle M, Sen PK: *Constrained Statistical Inference: Inequality.* Wiley, New York, NY: Order and Shape Restriction; 2005.