EDITOR'S CHOICE

A secure protocol for protecting the identity of providers when disclosing data for disease surveillance

Khaled El Emam,[1,2] Jun Hu,[3] Jay Mercer,[4] Liam Peyton,[3] Murat Kantarcioglu,[5] Bradley Malin,[6] David Buckeridge,[7] Saeed Samet,[1] Craig Earle[8]

► An additional appendix is published online only. To view this file please visit the journal online (www.jamia.org).

[1]Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada
[2]Paediatrics, University of Ottawa, Ottawa, Ontario, Canada
[3]School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, Canada
[4]Family Medicine, University of Ottawa, Ottawa, Ontario, Canada
[5]Computer Science, University of Texas at Dallas, Dallas, Texas, USA
[6]Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA
[7]Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada
[8]Institute for Clinical Evaluative Sciences and the Ontario Institute for Cancer Research, Toronto, Ontario, Canada

Correspondence to
Khaled El Emam, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada; kelemam@uottawa.ca

Received 16 January 2011
Accepted 3 February 2011

ABSTRACT

Background Providers have been reluctant to disclose patient data for public-health purposes. Even if patient privacy is ensured, the desire to protect provider confidentiality has been an important driver of this reluctance.

Methods Six requirements for a surveillance protocol were defined that satisfy the confidentiality needs of providers and ensure utility to public health. The authors developed a secure multi-party computation protocol using the Paillier cryptosystem to allow the disclosure of stratified case counts and denominators to meet these requirements. The authors evaluated the protocol in a simulated environment on its computation performance and ability to detect disease outbreak clusters.

Results Theoretical and empirical assessments demonstrate that all requirements are met by the protocol. A system implementing the protocol scales linearly in terms of computation time as the number of providers is increased. The absolute time to perform the computations was 12.5 s for data from 3000 practices. This is acceptable performance, given that the reporting would normally be done at 24 h intervals. The accuracy of detection disease outbreak cluster was unchanged compared with a non-secure distributed surveillance protocol, with an F-score higher than 0.92 for outbreaks involving 500 or more cases.

Conclusion The protocol and associated software provide a practical method for providers to disclose patient data for sentinel, syndromic or other indicator-based surveillance while protecting patient privacy and the identity of individual providers.

INTRODUCTION

Provider reporting of diseases to public-health authorities is common.[1 2] However, often there is under-reporting by physicians and hospitals, including for notifiable diseases and frequently by wide margins.[3–24] One causal factor for this under-reporting has been provider concerns about patient privacy.[8 9 11 13 15 19 21–23 25–28] Such a reluctance to disclose information has been noted in the past,[29 30] and exists despite the US Health Insurance Portability and Accountability Act Privacy Rule permitting disclosures of personal health information for public-health purposes without patient authorization.[27 29 31–34] Canadian privacy legislation in multiple jurisdictions also permits health-information custodians to disclose personal health information without consent for a broad array of public-health purposes, including chronic disease and syndromic surveillance.[35] Concerns about disclosing data are somewhat justified; however, as there have been documented breaches of patient data from public-health information custodians.[36–42]

One way to address patient privacy concerns is to de-identify the individual-level data before disclosure to public health, with the possibility of re-identification if an investigation or contact tracing is required.[31 42 43] However, even if patient privacy concerns are addressed, there have been other concerns about risks to physicians when patient information is disclosed,[13] and specifically disclosures for public-health purposes (unpublished data). At least five types of risks have been noted:

1. Legal exposure. Disclosures without individual patient consent have resulted in tortious or contractual claims of invasion of privacy, breach of confidentiality or implied statutory violations under state law,[44] and the increasing collection and disclosure of electronic information raises physicians' malpractice liability exposure.[45]
2. Compliance exposure. Physicians have concerns about information being used to evaluate compliance with clinical practice guidelines and compliance with pay for performance programs.[46] This concern increases with the amount of detail in the information that is collected.
3. Intrusive marketing. Providers do not want to be targeted by marketers who gain access to their patient information.[46 47]
4. Inference or disclosure of income data. Physicians and their professional associations consider the disclosure of income information a serious privacy breach.[46 48]
5. Inference or disclosure of performance or competitive data. It has been noted that '[h]ealthcare providers compete fiercely,' making it difficult to establish adequate trust for the exchange of health information among health information custodians.[49] Furthermore, some data sources for disease surveillance are proprietary, such that they may have reservations about data sharing. For example, schools may not want their absenteeism levels known to avoid political repercussions, and commercial pharmacies would be concerned about their sales data becoming known to potential investors and competitors.[27 50 51] Such custodians may not be willing to disclose information without their identity being masked.[50]

A distributed architecture for syndromic or other indicator-based surveillance can mask the identity of providers.[29 30 52–54] The sources provide count data to independent hubs; these data are aggregated and possibly analyzed by the hubs, and then forwarded to the public-health unit. However, the hubs need to be fully trusted to protect the identity of data sources. This means that if a hub is compromised, corrupted, or compelled to disclose information, then the raw data would reveal the identity of the data sources. Therefore, stronger protections than currently afforded by distributed architectures are needed to alleviate the data-sharing concerns noted above.

In this paper, we present a practical surveillance protocol for the secure multi-party computation of counts. The protocol also follows a distributed model, but it only requires semi-trusted parties. This protects the identity of data sources under different plausible threats. By addressing such concerns, we remove another barrier to the collection of data for disease surveillance.

METHODS

In the following narrative, we assume that the data sources are physician practices. This is for the purposes of illustration and ease of presentation. The descriptions, and our proposed protocol itself, would be applicable if the providers were, say, hospitals, pharmacies, or schools.

Trusted versus semi-trusted parties

With distributed surveillance protocols, individual practices send count data to hubs.[29] The hubs then aggregate the counts, perform additional analyses, and forward summaries or alerts to the public-health units. The hubs are considered trusted third parties because they will know the identity and counts of each practice. There are three challenges with having a trusted third party: (a) disclosures if a hub is compromised or corrupted, (b) compelled disclosures, and (c) all providers must trust the hub(s).

The first challenge is that if a hub's security is compromised, the adversary will have access to the identity of practices and their corresponding counts. A compromise can be due to either insiders or outsiders. A compromise can be as simple as a 'change your password' phishing attack to obtain the credentials of an employee of the hub. Many social-engineering techniques exist,[55–57] and have been used to obtain passwords and very personal information from individuals and organizations (as well as to commit more dramatic crimes such as bank robberies).[58 59] A recent review of data breaches indicated that 12% of data breach incidents involved deceit and social-engineering techniques.[60] Corruption can occur if an individual with access to the raw data within the hub is bribed or blackmailed to reveal information.

Second, a hub could be compelled to disclose personal health information, for example, in the context of litigation. For research, the National Institutes of Health can issue certificates of confidentiality to protect identifiable participant information from compelled disclosure, and allow researchers to refuse to disclose identifying information in any civil, criminal, administrative, legislative or other proceeding, whether at the federal, state or local level.[61] However, these would not be applicable to non-research projects or to projects that are not approved by an IRB, and most public-health surveillance programs would be in that excluded category. Furthermore, such certificates do not exist outside the USA.

Third, the hub must be trusted by all of the practices supplying data. This creates potential obstacles to the exchange of data across municipal, provincial/state, and international jurisdictions. To avoid sending data across jurisdictional boundaries, many regional hubs would need to be created. However, this will result in a proliferation of hubs and the replication of the exact infrastructure multiple times.

To address these challenges, we propose a distributed protocol with the weaker requirement of having only semi-trusted third parties. A semi-trusted third party would not be able to access any of the raw data, even if it wanted to. This means that if there is a security compromise, staff corruption, or a compelled disclosure, there is no additional risk of identifying practices. A protocol with semi-trusted third parties also overcomes the requirement of practices having to completely trust the hub. This allows us to set up a single infrastructure for a large number of practices across multiple jurisdictions. The only requirement on a semi-trusted third party is that it follow the protocol faithfully.

Context

The basic scenario we will use consists of the physician practices providing count data to a public-health unit. There are two types of counts disclosed over the reporting period: cases and all patients seen (denominators). We assume a 24 h reporting period, although our protocol would work with any interval, and that the counts are stratified by syndrome and age. The syndromes are influenza-like-illness (ILI) and gastrointestinal (GI). Ages are grouped similar to the CDC syndromic surveillance system[62] as <2, 2–4, 5–17, 18–27, 28–44, 45–64, 65+. Therefore, from each practice, we have a report containing 14 case counts for each age by syndrome stratum, and a total patient count for each age stratum for the previous 24 h. This makes a total of 21 counts per practice per reporting period.

Requirements for secure disease surveillance

The following are the requirements for a protocol that will allow meaningful reporting to a public-health unit while masking the identity of the reporting practices:

- R1. It should not be possible for any single adversarial party to know the true counts for any practice. This should hold even if a third party involved in the protocol is compromised, compelled to disclose its data, or corrupted.
- R2. The protocol should allow for technology failures. In a real-world setting, any distributed reporting system will have failures due to machine or connectivity breakdowns. The protocol should have inherent redundancy.
- R3. It must be possible to verify if a practice did submit data or did not submit data. This ensures data integrity and provides the basis for potentially compensating practices.
- R4. The computational requirements for the protocol should make it feasible to report at 24 h intervals.
- R5. It should be possible to identify practices with unusual spikes so that the public-health unit can obtain patient identities and initiate contact with them when necessary.
- R6. The ability to effectively detect disease outbreak clusters must not deteriorate with the secure protocol.

These requirements were constructed based on the authors' experiences and discussions with computer science and public-health professionals. They represent what are considered necessary conditions to protect the identity of patients and to allow public health to perform their surveillance and investigation functions effectively. The first four requirements address the trustworthiness of the protocol from the perspective of the patients, practices, and public-health units. The latter two requirements address the practical utility of the protocol to public health. Trustworthiness and practical utility are both

important, as they will increase the likelihood of initial adoption of the protocol and ongoing use.

Homomorphic cryptosystems

An important technique used in our protocol is homomorphic encryption. Utilizing a homomorphic cryptosystem, mathematical operations can be performed on the encrypted values (ciphertext) that produce the correct result when decrypted (plaintext). An example is additive homomorphic encryption introduced by Paillier,[63] in which, conceptually, the summation of two messages is equal to the decryption of the product of their corresponding ciphertexts:

$$D(E(m_1,e)\otimes E(m_2,e),d) = m_1 + m_2 \quad (1)$$

In this equation m_1 and m_2 are the two plaintext messages, E is the encryption function, D is the decryption function, e is the public encryption key, and d is the private decryption key. More details of the exact computation are provided in the appendix.

It is also possible to compute the product of a ciphertext with a constant q:

$$D(E(m_1,e)^q,d) = m_1 \times q \quad (2)$$

For example, if we want to convert the sign of a number, we would raise the power of the ciphertext to $q=-1$.

Another property of Paillier encryption is that it is probabilistic. This means that it uses randomness in its encryption algorithm so that when encrypting the same message several times, it will, in general, yield different ciphertexts. This property is important to ensure that an adversary would not be able to compare an encrypted message with all possible counts from zero onwards and ascertain the encrypted value.

A threshold version of the Paillier cryptosystem requires t out of l parties to decrypt a message.[64] For example, if we have a (2,3) threshold cryptosystem, we would need any two parties out of three to decrypt the message. No single party can decrypt the message.

Secure protocol for disease surveillance

The two phases of our secure protocol are illustrated in figures 1 and 2. Data would be aggregated into groups of at least k practices, where we set k=5 for illustrative purposes below. This means that it will not be possible for anyone but the practice itself to know the actual counts for any practice. The public-health unit will only be able to know the total count for groups of five practices or more.

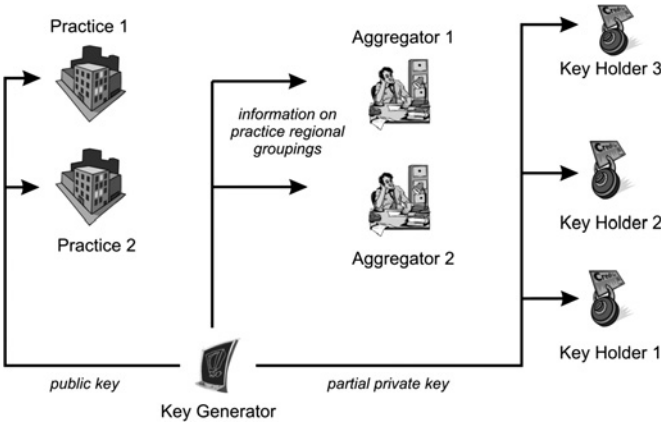

Figure 1 Set-up phase of the secure computation protocol assuming only two practices are submitting data.

Roles in our protocol

There are six roles in the protocol: (a) the practices—in the illustration we have only two practices, but this can be a much larger number; (b) the key generator (KG) issues the public and private keys for use by the various parties; (c) the aggregators are semi-trusted third-parties who perform the group- and stratum-specific sums of counts, (d) the key holders (KH) are semi-trusted third parties who decrypt the sums; (e) the mixer is a semitrusted third party who combines the results from at least two out of three key holders; and (f) the public-health unit (PHU) itself. A single physical entity can play multiple roles. An instance of a role will be referred to as a node.

The two aggregators are fully redundant in that the protocol can be implemented with a single aggregator. The primary purpose for redundancy is to ensure that the aggregation operations are performed even if a single aggregator fails or is not accessible. There are three KHs to ensure that there is redundancy built into the system. This means that any single KH can fail, but the overall results can still be computed. For additional robustness, it is also possible to extend the protocol to have $t>2$ and/or $l>3$ key holders (eg, 2 out of 4) to be able to decrypt the counts. A minimalist implementation of the protocol with no redundancy would have only one aggregator and two KHs.

The exact configuration in figures 1 and 2 is the one we have used in our demonstration system. The protocol has two main phases described below.

Set-up phase of our protocol

At the outset, the KG generates a public key and the corresponding partial private keys. The public key is given to all of the participating practices when they register. The partial private keys are sent to each of the three KHs. The KG then destroys its copies of the partial private keys after their successful transmission.

During set-up, each practice registers with the KG to indicate that it wishes to participate in the protocol. Registration means downloading the client software and a configuration file. When a practice registers, they also provide a physical address which can be used to identify the other geographically closest registered practices. The configuration file contains the public key as well as the regional grouping of the closest registered practices. The provider installs the client software and is ready to submit count data. The KG informs the aggregators about each practice that has registered and its regional grouping.

For the sake of example, we will assume that we have two regional groupings of practices: Ottawa and Montreal. More formally:

- Assume there are P strata, M practices, N aggregators, T KHs, and R regional groups. In our example, we have 21 strata, two aggregators, three KHs, and two groups.
- KG generates the public key PK and the T partial private keys SK_t where $t \in \{1 \ldots T\}$, and sends PK to each new practice when it joins the protocol, and sends the SK_t to each of the KHs.
- Each new practice has a unique ID. All aggregators are informed of each practice's unique ID and group when they join. The PHU is informed of all practices within a group.

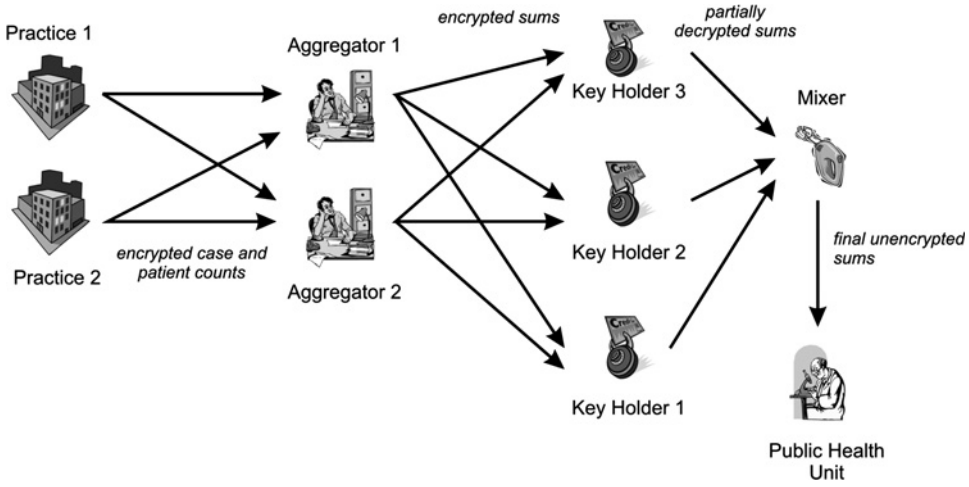

Figure 2 Actual operation of the protocol to securely compute counts.

Operation phase of our protocol

At the end of the 24 h period, each practice computes the counts for the 21 strata, each of which is encrypted using PK. The encrypted counts are sent by each practice to all of the aggregators. Within each group, an aggregator will sum the encrypted counts only if they come from at least k practices. For example, if the Ottawa region only submitted the counts from four practices, the aggregator would produce a 'NO DATA' result for Ottawa.

If at least k encrypted values for a group have been received by an aggregator, the aggregator computes the sums within each stratum across all of the practices within each group. The aggregator does not know what the original values from each practice are, and does not know what the sums are because all of these values are encrypted. The aggregator then sends the encrypted P group sums for the R groups to each KH (except for the groups with 'NO DATA' status).

Each KH uses its partial private key to decrypt the sums it receives, which are subsequently sent to the mixer. The KHs ignore regions with no data.

The Mixer selects any two KH values and computes the decrypted values of the P group sums for the R groups with data, which is forwarded to the PHU. More formally:

1. Each practice computes E_{ij} where $i \in \{1,\ldots,P\}$ and $j \in \{1,\ldots,M\}$ as: $E_{ij}=E(C_{ij}, PK)$ where C_{ij} is the count for stratum i for practice j.
2. The $P \times E_{ij}$ values for each practice are then sent to all of the aggregators.
3. Each aggregator sums (which is equivalent to a multiplication of encrypted values as in equation 1) the values within each stratum within each group: $S_{ir} = \prod_{j \text{ in group } r} E_{ij}$ where $r \in \{1,\ldots, R\}$.
4. The sums are sent by the aggregator to each of the KHs. The KHs decrypt the sums using their partial decryption key: $s_{irt}=D(S_{ir}, SK_t)$, which are sent along with their validity proofs to the mixer.
5. The mixer verifies the partial KH decryptions using their proofs (see the appendix). It then selects any two valid decryption results and combines their results to obtain the final count: s_{ir}.

At the end of these steps, the PHU has the plaintext counts for each group for each stratum.

MEETING THE REQUIREMENTS

Security analysis (R1)

The security analysis for this protocol is provided in the appendix. This demonstrates that no party will know the practice identities and their counts under plausible compromises or corruption of individual nodes and collusions among nodes.

Node failures (R2)

Our protocol has multiple points of redundancy, making allowances for real-world failures of nodes. This meets requirement number 2. A simulation of node failures is presented in the appendix. This demonstrates that with two aggregators, if any aggregator has a failure rate as high as 20%, there will still be at least one aggregator operating around 98% of the time. With a KH node failure rate of 15%, at least two nodes will be operating around 95% of the time.

Detecting practices providing and not providing data (R3)

An important element of a real-world deployment is the use of digital signatures. Digital signatures will ensure that the senders of messages are who they say they are (authenticity), that the messages cannot be modified in transit without the tampering being discovered (integrity), and that the senders cannot claim that they did not send the messages (non-repudiation). Digital signatures will make it possible to ensure that data are indeed coming from the practices and that practices cannot deny that they submitted counts. Digital signatures and their application in our protocol are described in the appendix.

There will be situations when the PHU needs to detect if any practices are consistently not providing data but claiming that they are. In a sense, the PHU needs to detect 'free riding' practices whose lack of contribution of counts is hidden within the practice group total. Such free riding may be deliberate or accidental. For example, a practice may insist that it is providing data and that the aggregator is 'losing it.' A proof of data submission would be particularly important if practices are compensated financially for providing data. In such a case, the PHU would need to verify which practices have been contributing counts. For example, if there were eight practices in a group and only five provided data, and the remaining three insist that their systems are working and sending data, the PHU can verify whether the three missing practices did indeed provide their counts. In the appendix, we provide an extension to the protocol that can be used by the PHU to verify which practices have provided data in their group. The approach checks membership using a commutative hash function,[65] and makes it impossible for a practice to misrepresent that it provided a count in a total.

Computation performance (R4)

A critical concern with protocols utilizing secure multi-party computation is their performance under realistic situations. We

conducted a performance evaluation of the surveillance protocol to determine how it scales as the number of practices and groups increases. Performance is defined in terms of the amount of communication and time taken to perform the computations. The assessment in the appendix shows that only a handful of messages need to be communicated among nodes, and the absolute time to perform the computations was 12.5 s for data from 3000 practices.

Contacting patients (R5)

If full identifying information about cases is sent to the PHU, it can contact the patients directly if an investigation needs to be initiated. However, under our protocol, only count data about patient encounters are sent to the PHU. For sentinel surveillance programs, this is generally not problematic because contacting patients is not usually done. However, for other types of indicator-based surveillance, the PHU may want to contact patients under certain circumstances.

The PHU would first need to find out which practices have unusually high counts that require investigation. Subsequently, these practices are contacted and asked to identify the cases. Each practice has access to a line list of the individual level records that make up each stratum count, and therefore can respond with more detailed information about specific patients. The PHU only needs to determine which practice(s) have unusual spikes.

We present a protocol extension in the appendix for identifying the N practices with the largest counts within a regional group. This protocol does not reveal the actual counts from any of these practices, only that they have the largest counts in their group. The PHU can then identify the practice(s) with the highest counts and contact them for additional details. These details could include detailed line listings of the patients who made up specific strata.

Detecting disease outbreaks (R6)

The ability to detect spatial clusters is important for disease surveillance. We consider two scenarios.

For the first scenario, the counts are stratified by some geographic area, such as the census tract. This would indicate the counts of patients in each area aggregated across practices. It would be necessary to ensure that the areas are large enough to protect patient identifiability, however.[66 67] Since our protocol would not affect these counts, the ability to detect clusters will be the same as current distributed surveillance protocols.

In the second scenario, the strata sent to the PHU do not contain patient-specific location information. Therefore, the PHU could perform clustering on the practices themselves to detect geographically adjacent practices with unusually high cases. The question is whether the grouping of practices masks the ability to detect such practice clusters. In the appendix, we present the results of a simulation demonstrating that, for practice groups of size 5 and 10, the accuracy of cluster detection is quite high (F-scores greater than 0.95 and 0.92 respectively) and similar to when the practices are not grouped.

DEPLOYMENT CONSIDERATIONS

For deployment, two aggregators and KH pairs can coexist in the same physical node/site, since collusion between them would not reveal any new information. In addition, the mixer and the public-health unit can exist on the same physical node/site. This is illustrated in figure 3, which shows that only four nodes/sites would be needed to implement the protocol as described. The same nodes can support multiple local and national surveillance initiatives. Additional practical deployment considerations are provided in the appendix.

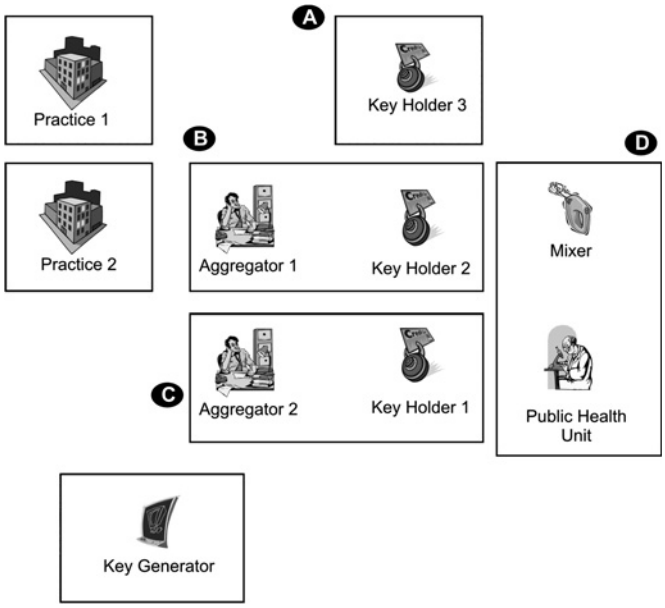

Figure 3 Minimalist deployment of the protocol. Each box represents a single node. Note that there may be multiple practice nodes; we have shown only two here.

It would be important to convince providers to participate in such a surveillance protocol. A recent study examining family doctor attitudes toward the disclosure of patient data for public-health purposes determined that an endorsement by their professional college would be a key factor in their willingness to participate in disclosures of data to public health (unpublished data). The reasoning would be that the college would be an independent and trusted party that would provide an objective opinion regarding the trustworthiness of the protocol. Therefore, as an initial step for implementation, it will be important to engage with the professional colleges and work with them to transition such a protocol into practice.

Acknowledgments We wish to thank P AbdelMalik, S Rose and G Middleton for their help in the data preparation and spatial analysis for the practice aggregation problem.

Funding This work was partially funded by the Canadian Institutes of Health Research, the GeoConnections program of Natural Resources Canada, the Ontario Institute for Cancer Research, the Natural Sciences and Engineering Research Council and grant number R01-LM009989 from the National Library of Medicine, National Institutes of Health.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Chorba TL,** Berkelman RL, Safford SK, *et al*. Mandatory reporting of infectious diseases by clinicians. *JAMA* 1989;**262**:3016–26.
2. **Roush S,** Birkhead G, Koo D, *et al*. Mandatory reporting of diseases and conditions by health care professionals and laboratories. *JAMA* 1999;**281**:164–70.
3. **Standaert SM,** Lefkowitz LB Jr, Horan JM, *et al*. The reporting of communicable diseases: a controlled study of *Neisseria meningitides* and *Haemophilus influenzae* infections. *Clin Infect Dis* 1995;**20**:30–6.
4. **Doyle TJ,** Glynn MK, Groseclose SL. Completeness of notifiable infectious disease reporting in the United States: an analytical literature review. *Am J Epidemiol* 2002;**155**:866–74.
5. **Watkins M,** Lapham S, Hoy W. Use of medical center's computerized health care database for notifiable disease surveillance. *Am J Public Health* 1991;**81**:637–9.
6. **Thacker SB,** Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev* 1988;**10**:164–90.

7. **Szecsenyi J,** Uphoff H, Ley S, *et al*. Influenza surveillance: experiences from establishing a sentinel surveillance system in Germany. *J Epidemiol Community Health* 1995;**49**(Suppl 1):9—13.
8. **Cleere RL,** Dougherty WJ, Fiumara NJ, *et al*. Physician's attitudes toward venereal disease reporting. A survey by the National Opinion Research Center. *JAMA* 1967;**202**:117—22.
9. **Rushworth RL,** Bell SM, Rubin GL, *et al*. Improving surveillance of infectious disease in New South Wales. *Med J Aust* 1991;**154**:828—31.
10. **Calzavara LM,** Coates RA, Craib KJ, *et al*. Underreporting of AIDS cases in Canada: a record linkage study. *CMAJ* 1989;**142**:36—9.
11. **Johnson RJ,** Montano BL, Wallace EM. Using death certificates to estimate the completeness of AIDS case reporting in Ontario in 1985—87. *CMAJ* 1989;**141**:537—40.
12. **Jenkinson D.** Whooping cough: what proportion of cases is notified in an epidemic? *BMJ (Clin Res Ed)* 1983;**287**:185—6.
13. **Jones J,** Meyer P, Garrison C, *et al*. Practitioner HIV/AIDS reporting characteristics. *Am J Public Health* 1992;**82**:889—91.
14. **Bernillon P,** Lievre L, Pillonel J, *et al*. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990—1993. The Clinical Epidemiology Group from Centres d'Information et de Soins de l'Immunodéficience Humaine. *Int J Epidemiol* 2000;**29**:168—74.
15. **Konowitz PM,** Petrossian GA, Rose DN. The underreporting of disease and physicians' knowledge of reporting requirements. *Public Health Rep* 1984;**99**:31—5.
16. **Schramm MM,** Vogt RL, Mamolen M. The surveillance of communicable disease in Vermont: who reports? *Public Health Rep* 1991;**106**:95—7.
17. **Clarkson JA,** Fine PE. The efficiency of measles and pertussis notification in England and Wales. *Int J Epidemiol* 1985;**14**:153—68.
18. **Macleod CA.** Haemophilus influenzae: the efficiency of reporting invasive disease in England and Wales. *Commun Dis Rep CDR Rev* 1994;**4**:R13—16.
19. **Rothenberg R,** Bross DC, Vernon TM. Reporting of gonorrhea by private physicians: a behavioral study. *Am J Public Health* 1980;**70**:983—6.
20. **Rivest P,** Sagot B, Bedard L. Evaluation of the completeness of reporting of invasive meningococcal disease. *Can J Public Health* 1999;**90**:250—2.
21. **Scatliff JN.** Survey of venereal disease treated by Manitoba physicians in 1972. *Can Med Assoc J* 1974;**110**:179—82, passim.
22. **Marier R.** The reporting of communicable diseases. *Am J Epidemiol* 1977;**105**:587—90.
23. **Gelman AC,** Vandow JE, Sobel N. Current status of venereal disease in New York City: a survey of 6,649 physicians in solo practice. *Am J Public Health Nations Health* 1963;**53**:1903—18.
24. **Allen CJ,** Ferson MJ. Notification of infectious diseases by general practitioners: a quantitative and qualitative study. *Med J Aust* 2000;**172**:325—8.
25. **AbdelMalik P,** Boulos MN, Jones R. The perceived impact of location privacy: a web-based survey of public health perspectives and requirements in the UK and Canada. *BMC Public Health* 2008;**8**:156.
26. **Drociuk D,** Gibson J, Hodge J Jr. Health information privacy and syndromic surveillance systems. *Morb Mortal Wkly Rep* 2004;(53 Suppl):221—5.
27. **Wojcik R,** Hauenstein L, Sniegoski C, *et al*. Obtaining the data. In: Lombardo J, Buckeridge D, eds. Disease surveillance: a public health informatics approach. John Wiley & Sons, Hoboken, NJ, USA, 2007.
28. **Hodge JG Jr.** Health information privacy and public health. *J Law Med Ethics* 2003;**31**:663—71.
29. **Lazarus R,** Yih K, Platt R. Distributed data processing for public health surveillance. *BMC Public Health* 2006;**6**:235.
30. **Sengupta S,** Calman NS, Hripcsak G. A model for expanded public health reporting in the context of HIPAA. *J Am Med Inform Assoc* 2008;**15**:569—74.
31. **Broome CV,** Horton HH, Tress D, *et al*. Statutory basis for public health reporting beyodn specific diseases. *J Urban Health* 2003;**80**(2 Supp 1):I14—22.
32. **Centers for Disease Control and Prevention (CDC).** HIPAA privacy rule and public health. Guidance from CDC and the US Department of Health and Human Services. *MMWR Morb Mortal Wkly Rep* 2003;(52 Suppl):1—17, 19—20.
33. **Gostin L.** *Public Health Law*. Berkeley, CA: University of California Press, 2008.
34. **El Emam K,** Fineberg A. *Risk Assessment for the Disclosure of Personal Health Information for Public Health Purposes*. Ottawa, Canada: Public Health Agency of Canada, 2009.
35. **Landry S.** AIDS list is out: State investigating breach. In: St Petersburgh Times, 20 Sept 1996: 1-A, 10-A.
36. **Vyhnak C.** Health records of thousands lost in Durham. The Toronto Star, 2009. http://www.thestar.com/news/gta/article/741816—health-records-of-thousands-lost-in-durham. http://www.webcitation.org/5qlvIluDT.
37. **Detroit Health Department.** *Records Missing: Computers Stolen From Medical Building*. clickondetroit.com. 2009. http://www.clickondetroit.com/news/21973152/detail.html. http://www.webcitation.org/5qlvhkQ7o.
38. Kanawha-Charleston Health Department. datalossdb. http://datalossdb.org/incidents/1519. http://www.webcitation.org/5qlvwwtzt.
39. **McDougall P.** Lost disc puts 2.9 million georgia residents at risk For ID theft informationweek, 2007. http://www.informationweek.com/news/security/showArticle.jhtml?articleID=198900446. http://www.webcitation.org/5qlw76moO.
40. **Canalis J.** Lowenthal offers ID theft help after state's Social Security foul-up. Press-Telegram, 2010. http://www.presstelegram.com/news/ci_14413495. http://www.webcitation.org/5qlxP2JgW.
41. *Patients Not Notified That Their Health Records Were Stolen: Information Was Being Collected For National Autism Study*. The Denver Channel, 2005. http://www.thedenverchannel.com/7newsinvestigates/4438964/detail.html. http://www.webcitation.org/5qlxfrqWK.
42. **Birtwhistle R,** Keshavjee K, Lambert-Lanning A, *et al*. Building a pan-Canadian primary care sentinel surveillance network: initial development and moving forward. *J Am Board Fam Med* 2009;**22**:412—22.
43. **Wong J,** Mercer J, Nizar SM, *et al*. Rapid real time surveillance and monitoring of pandemic influenza associated pneumonia & risk factors using primary care electronic medical records (EMR). *14th International Congress on Infectious Diseases (ICID)*, 2010.
44. **Hodge JG Jr,** Gostin LO, Jacobson PD. Legal issues concerning electronic health information: privacy, quality, and liability. *JAMA* 1999;**282**:1466—71.
45. **Mangalmurti SS,** Murtagh L, Mello MM. Medical malpractice liability in the age of electronic health records. *N Engl J Med* 2010;**363**:2060—7.
46. **Field RI.** Physician rights to privacy of data prevail in two major court tests but new questions lie ahead. *Healthcare and Law* 2009;**34**:193—5.
47. **Kosseim P,** El Emam K. Privacy interests in prescription records, part 1: prescriber privacy. *IEEE Security and Privacy* 2009;**7**:72—6.
48. **CBC News.** Kennedy brushes off privacy breach accusation CBC, 2010. http://www.cbc.ca/canada/ottawa/story/2010/11/09/kennedy-privacy-breach-119.html. http://www.webcitation.org/5uEhA3v6X.
49. **Gravely SD,** Whaley ES. The next step in health data exchanges: trust and privacy in exchange networks. *J Healthc Inf Manag* 2009;**23**:33—7.
50. **Kress A.** Data provider relationships: pros, cons, and considerations. *Morb Mortal Wkly Rep* 2004;**53**(Suppl):247.
51. **Lober WB,** Trigg L, Karras B. Information system architectures for syndromic surveillance. *MMWR Morb Mortal Wkly Rep* 2004;**53**(Suppl):203—8.
52. **Lazarus R,** Klompas M, Campion FX, *et al*. Electronic support for public health: validated case finding and reporting for notifiable diseases using electronic medical data. *J Am Med Inform Assoc* 2009;**16**:18—24.
53. **Platt R,** Bocchino C, Caldwell B, *et al*. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. *J Urban Health* 2003;**80**(2 Supp 1): I25—31.
54. **Yih WK,** Cadwell B, Harmon R, *et al*. National bioterrorism syndromic surveillance demonstration program. *Morb Mortal Wkly Rep* 2004;**53**(Suppl):43—9.
55. **Dolan A.** *Social Engineering*. Bethesda, MD: SANS Institute, 2004.
56. **Centre NISC.** *Social Engineering Against Information Systems: What is it and How do you Protect Yourself?* 2006.
57. **Winkler I,** Dealy B. Information security technology ? Don't rely on it: A case study in social engineering. *Proceedings of the Fifth USENIX UNIX Security Symposium*, 1995.
58. **Mitnick K,** Simon W. *The Art of Deception: Controlling the Human Element of Security*. Indianapolis: Wiley, 2002.
59. **Long J.** *No Tech Hacking: A Guide to Social Engineering, Dumpster Diving, and Shoulder Surfing*. Burlington, MA: Syngress Publishing Inc, 2008.
60. *Verizon Business Risk Team. 2009 Data Breach Investigations Report*, 2009.
61. **Department of Health and Human Services.** *Certificates of Confidentilaity Kiosk*. http://grants.nih.gov/grants/policy/coc/.
62. **Centers for Disease Control.** *Aggregate Data Use and Sharing Agreement for Novel H1N1 Influenza Surveillance in 2009—2010*, 2009.
63. **Paillier P.** *Public-Key Cryptosystems Based on Composite Degree Residuosity Classes*. Prague: EUROCRYPT'99, 1999.
64. **Fouque PA,** Poupard G, Stern J. Sharing decryption in the context of voting or lotteries. *Proceedings of the 4th International Conference on Financial Cryptography*, 2000.
65. **Benaloh J,** de Marc M. *One-Way Accumulators: A Decentralized Alternative to Digital Signatures*. Perugia: Advances in Cryptography (EUROCRYPT), 1994.
66. **El Emam K,** Brown A, AbdelMalik P, *et al*. A method for managing re-identification risk from small geographic areas in Canada. *BMC Med Inform Decis Mak* 2010;**10**:18.
67. **El Emam K,** Brown A, AbdelMalik P. Evaluating predictors of geographic area population size cutoffs to manage re-identification risk. *J Am Med Inform Assoc* 2009;**16**:256—66.