Challenging the assertion of comparability of surveillance and administrative data

Tara Leigh Donovan MSc[1], Leslie Forrester MSc[2], Jun Chen Collet MSc[1], Louis Wong MPH[3], Julie Mori PhD[4], Elisa Lloyd-Smith PhD[2], Blair Ranns MPH[5] and Guanghong Han PhD[1]

[1]Provincial Health Services Authority, British Columbia, Canada, [2]Vancouver Coastal Health Authority, British Columbia, Canada, [3]Mount Sinai Hospital, Toronto, Ontario, Canada, [4]Interior Health Authority, British Columbia, Canada and [5]Island Health Authority, British Columbia, Canada

To the Editor—In the April 2017 issue of *Infection Control and Hospital Epidemiology*, we read with interest the study by Ramirez Mendoza et al[1] that compared administrative versus surveillance data for capturing hospital-associated methicillin-resistant *Staphylococcus aureus* (MRSA) infections in Canadian hospitals. However, due to the lack of a clear objective and some methodological concerns, we disagree with their study conclusion that there is good evidence for comparability between administrative and surveillance data.

The primary objective stated in the abstract was to assess the *accuracy* of administrative data concerning in-hospital bloodstream infections (BSIs) and all-body-site infections due to MRSA. However, in the body of the paper, the last paragraph of the background mentions that the primary objective of the study was to assess the *feasibility* of using administrative data to monitor these MRSA infections. A secondary objective was to determine whether there was a *correlation* between MRSA BSIs and all-body-site MRSA infections within the administrative data only. This lack of clarity around the primary objective is a concern because the methodological approach to determine feasibility differs from establishing accuracy.

If the author's intention was to assess accuracy, the high correlation coefficient (r) reported in the paper does not necessarily mean high agreement between 2 datasets, nor accuracy of the data. For example, if the number of cases counted is exactly the same between 2 datasets, the cases themselves may not represent the same patients. Thus, a strong correlation between datasets is not meaningful unless the comparison of the datasets was conducted using patient-level data, which this study did not. Moreover, if the intention was to assess reliability of the datasets, we believe it would have been more appropriate to perform Lin's concordance correlation coefficient.[2] Lin's coefficient adjusts the Pearson correlation coefficient by assessing both how close the data are about the line of best fit and the precision of agreement.

The administrative data from Alberta identified 113 MRSA infections, for a rate of 0.43 per 10,000 patient days, compared to 229 cases and a rate of 0.90 MRSA infections identified by surveillance data. Although the correlation coefficient indicated comparability (r = 0.92; 95% CI, 0.88–0.94; $P < .0001$), the rate was twice the amount, meaning that for every 1 case in the administrative dataset there were 2 cases in the surveillance

dataset. Furthermore, there may be intrinsic differences in the cases identified that impact the ability to use administrative data in place of surveillance data. Finally, conducting a comparison of this data using different infection sources (ie, MRSA bacteremia BSI in the Ontario dataset and all-body-site MRSA infections in the Alberta dataset) is of vital concern considering the differences between these datasets and the noted challenges with using administrative data to identify infections.

We challenge the assertion of comparability of surveillance and administrative data based on how the data are generated. Surveillance data employ comprehensive definitions and case finding methods, which includes a review of positive laboratory results and clinical data (ie, chart reviews, in-person interviews, etc) to capture hospital infections. Administrative data are based on applying *International Classification of Disease, Tenth Revision, Canada* (ICD-10-CA) diagnostic codes against information documented in the patient chart. The quality and completeness of clinical documentation are recognized limitations of administrative data; they impact coders' ability to accurately code cases, potentially leading to an overestimate or underestimate of cases. Correctly identifying a patient with an infection based on coding alone is challenging because it can be difficult to differentiate an infection from colonization for nonsterile sites. In a study involving professional coders in Alberta, Tang et al[4] identified multiple barriers to producing high-quality administrative data, including documentation that is incomplete and nonspecific and often contains errors and discrepancies.[3] Furthermore, Nicholls et al[4] emphasized the importance of validating coding accuracy to reduce the potential of misclassification bias.

This study used data from Alberta and Ontario, and we do not think it is appropriate to generalize the conclusion for Canadian hospitals. Two unpublished analyses from 2 health regions in British Columbia were not in agreement with this study conclusion. An analysis conducted by Interior Health compared MRSA in-hospital infections from the Discharge Abstracts Database (DAD) to new healthcare-associated MRSA infection cases from the surveillance database of 22 acute-care facilities. The Lin's concordance coefficient was 0.31.[5] For the same period, a patient-level record-linkage analysis involving 9 acute-care facilities conducted by Vancouver Coastal Health found that the DAD data only captured 30% of the 180 healthcare-associated MRSA infection cases identified through surveillance. Both analyses demonstrate the poor concordance between administrative and surveillance data for identifying in-hospital MRSA infections.

In conclusion, we cannot support the authors claim that there is good evidence of comparability of administrative and

surveillance data in Canadian hospitals based on problematic data and methods used.

References

1. Ramirez Mendoza JY, Daneman N, Elias MN, Amuah JE, Buth K, Couris CM et al. A comparison of administrative data versus surveillance data for hospital-associated methicillin-resistant *Staphylococcus aureus* infections in Canadian hospital. *Infect Control Hosp Epidemiol* 2017;38: 436–443.
2. Watson FP, Petrie A. Method agreement analysis: a review of correct methodology. *Theriogenology* 2010;73:1167–1179.
3. Tang KL, Lucyk K, Quan H. Coder perspectives on physician-related barriers to producing high-quality administrative data: a qualitative study. *CMAJ* 2017;5:E617–E621.
4. Nicholls SG, Langan SM, Benchimol EI. Routinely collected data: the importance of high-quality diagnostic coding to research. *CMAJ* 2017 August 21;189:E1054–E1055.
5. Lin's concordance correlation coefficient. Marta Garcia-Granero SPSS Macros. https://gjyp.nl/marta/. Published 2010. Accessed April 2017.

Going back to prospectively collected results with a probiotic for primary prevention of *Clostridium difficile* infection at a tertiary-care medical center

Noam Ship PhD and Serge Carrière MD, FRCP

Research and Development, Bio-K Plus International, Laval, Quebec, Canada

To the Editor—Implementing a facility-level *Clostridium difficile* infection (CDI) prevention protocol is a challenging endeavor, and Trick et al[1] should be commended for their early successes at a large institution. As the manufacturers of the probiotic comprising *L. acidophilus* CL1285, *L. casei* LBC80R, and *L. rhamnosus* CLR2 (Bio-K+), we contributed our products to 9,072 patients for this study at no cost to the investigators as well as monetary support for a research assistant to collect data. We watched the evolution of this quality improvement study in anticipation, receiving regular updates on recruitment and product consumption, though we had no role in the collection or interpretation of the data.

Several issues emerged that made the investigators "unable to electronically extract patient-level antibiotic and probiotic receipt data." In the absence of a mechanism to review the full dataset of thousands of eligible patients, the authors undertook a case-control study to examine 68 cases and 68 matched controls in detail using a retrospective chart review. One conclusion from this study was that after adjusting for severity of illness, and temporal and spatial proximity, there was "no protective effect from the probiotic." It is our opinion that the case-control study was rigorous in considering the available data. Many of the known risk factors were controlled, but the principal modifiable risk factor for CDI—antibiotics—was mostly overlooked. Tartof et al[2] elegantly showed a 2-fold increase in the risk of CDI with each additional antibiotic administered. Among 401,234 adults admitted to Kaiser Permanente Southern California hospitals, 0.5% tested CDI-positive when taking 1 antibiotic, 1.0% when taking 2 antibiotics, and 2.3% when taking 3 or more antibiotics.

It may be, for example, that the composition of a patient's intestinal microbiota prior to antibiotic exposure is another relevant predictor of susceptibility to *C. difficile* overgrowth and infection. A detailed metagenomics analysis by Raymond et al[3] of stool samples from healthy volunteers taking an antibiotic uncovered consistent impairments to the diversity and functioning of the gut microbiota and enrichment of resistance genes. In addition, the initial microbiome composition among certain volunteers predicted an overgrowth of known pathogens. Applied to the hospital setting, it could be that hospitalized patients who develop a CDI are inherently more susceptible to the effect of antibiotics on their microbiome. Scientific questions like these are beyond the technical limitations of the case-control design.

The case-control study did not detect a protective effect, but results from the same cohort presented in 2015 indicate a reduced risk of CDI in patients treated with this probiotic.[4] A risk ratio of 0.6 (95% CI, 0.4–0.9; $P = .01$) was calculated, representing a significant protective effect from *C. difficile* infection for the probiotic recipients (Fig. 1). The inputs for this equation are based on the principal data set collected prospectively within the design of the study, and as such, they do not rely on detailed electronic patient records. Thus, these prospectively collected data suggest that the intervention was having the intended effect.

Although Trick et al describe challenges in implementing the protocol, in following patients, and in extracting patient-level data, this quality improvement study provides invaluable information on the real-world practical application of probiotics in the fight against *C. difficile* infection. As noted in a public release by the Society for Healthcare Epidemiology of America (SHEA), it is critically important to test emerging interventions in routine practice and to learn from those