Appendix 1 (as supplied by the authors): Development and validation of an algorithm to identify pertussis immunization using Ontario Health Insurance Plan physician billing claims

Algorithm development

The Ontario Health Insurance Plan (OHIP) database contains physician billing information on virtually all Ontario residents starting 1 January 1992, except for new residents within the previous 3 months.[1] OHIP introduced vaccine-specific immunization billing claims in 2011, replacing the previously used general immunization codes.

We extracted general (immunization with [G538] and without [G539] a consultation) and pertussis vaccine-specific codes (G841 [Diphtheria-Tetanus-acellular Pertussis-Inactivated Polio-*Haemophilus influenza* type B; DTaP-IPV-Hib], G840 [Diphtheria-Tetanus-acellular Pertussis-Inactivated Polio; DTaP-IPV], G847 [Tetanus-diphtheria-acellular pertussis; Tdap]) for all individuals in the Public Health Ontario laboratory database of pertussis tests. This resulted in 29,300 entries from 7,560 subjects, including 2,073 unvaccinated individuals (Figure e1). Since 95% were general immunization codes (G538/G539), we felt it was necessary to 'clean' the dataset to extract immunizations likely to represent a pertussis-containing vaccine, and remove others (e.g., pneumococcal, meningococcal, measles-mumps-rubella, varicella). We developed an algorithm based on the 2013 Ontario immunization schedule.[2] In summary, the rules were:

1. The first pertussis vaccine cannot be administered earlier than six weeks of age.
2. The second and third doses cannot be administered sooner than 28 days following the previous dose.

Appendix to: Schwartz KL, Kwong JC, Deeks SL, et al. Effectiveness of pertussis vaccination and duration of immunity.
CMAJ 2016. DOI:10.1503/cmaj.160193.

3. A child cannot receive more than three pertussis vaccines during the first 12 months of life.
4. The fourth dose cannot be administered earlier than 17 months of age.
5. A child cannot receive more than one pertussis booster between 17 months and four years of age.
6. The fifth dose cannot be administered before four years of age.
7. A child cannot receive their sixth dose until 13.5 years of age.
8. A child cannot receive more than one booster beyond 13.5 years of age.
9. Catch-up doses were permitted up to the maximum number of doses expected for the age of the child.

Figure e1 details the number of vaccines removed at each step of the algorithm. In total, 8,663 (32%) of the general immunization codes were removed, leaving 18,554 immunizations given to 5,487 individuals (totaling 7,560 including the 2,073 unvaccinated individuals, many of whom were infants less than three months of age).

Algorithm validation

We validated the above algorithm against the reference standard of documentation in a primary care electronic medical record (EMR) system. The methodology was similar to a recent study validating OHIP billing claims for infant immunizations.[3]

We included birth cohorts of Ontario children under the care of family physicians who share their practices' EMR data with the Electronic Medical Record Administrative data Linked Database (EMRALD), a centralized repository of EMR data used for research and evaluation.[4]

EMRALD comprises EMR data from over 350 Ontario family physicians who use Practice Solutions Suite® software, the most widely used EMR in Ontario.[5] Individual-level data from EMRALD are collected semi-annually and linked to other administrative databases at the Institute for Clinical and Evaluative Sciences (ICES). Information collected in EMRALD includes; clinical encounters, the cumulative patient profile, family history, allergies, immunizations, diagnostic tests, prescriptions, discharge summaries, and consultation notes. Participating physicians contribute to EMRALD on a voluntary basis. EMRALD has been previously used for administrative data validation studies to identify patients with ischemic heart disease, stroke, and epilepsy.[6-8]

Children born between 1 January 2002 and 31 August 2004 and followed continuously for seven years in EMRALD were included for analysis. We excluded children who had duplicate identification numbers (preventing 1:1 linkage between EMRALD and the OHIP database) and those who died during the observation period. We also excluded children with no billed primary care assessments, and those with fewer than four visits to their family physician during the first year of life. These children were likely to have resided in Ontario intermittently or received their primary care from a salaried physician who did not submit billing claims to OHIP.

We conducted text searches for each vaccine of interest in the prescriptions and immunizations fields in EMRALD using a series of keywords (e.g., DTaP, Diptheria-Tetanus-acellular pertussis, pentacel, quadracel, adacel, etc). We also searched EMRALD for records of immunization billing claims submitted through the EMR software.

We then extracted OHIP billing claims for vaccination using the codes listed above for our study cohort, and applied the algorithm described above.

Using EMRALD immunization status as the reference standard, we calculated performance measures (i.e., sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) with 95% confidence intervals for the above algorithm at five intervals based on the age expected to have received the first, second, third, fourth, and fifth doses (i.e., three months, five months, seven months, 19 months, and seven years, respectively).

We identified 1,511 children followed in EMRALD for a minimum of seven years and linked to the OHIP database. The performance of the algorithm was quite good in infants younger than 12 months of age as sensitivity was 88%, 83%, and 79% at three, five, and seven months respectively (Table e1). However, sensitivity decreased with age to 66% at 19 months, and 57% at seven years. Specificity was 95% or higher from three to 19 months, and 83% at seven years. PPV was very high at 95% or higher at all ages, and NPV was generally low ranging from 25% to 65%. OHIP had a sensitivity of 67% and specificity of near 100% to identify completely unvaccinated children at seven years of age. However, the confidence intervals were wide due to the small number of unvaccinated children.

In summary, this pertussis immunization algorithm had high specificity for identifying true pertussis immunizations in children younger than seven years of age. The sensitivity in infants was also high but decreased with age. The reasons for this decline are not entirely clear. Physician remuneration for providing vaccines is small and ranges from $0.68 to $4.50 depending on the payment structure of the practice. It is possible that nurses or other non-billing providers may have administered vaccines more commonly in older children resulting in under-billing. A separate validation study using similar methods to evaluate vaccine-specific and general immunization codes in infants found similar results in this population.[3]

Important limitations of using an EMR as the reference standard for immunizations should be noted. EMR records are dependent on the clinician documenting administration of the vaccine. There are multiple areas of free text within the EMR, and while we attempted to capture all relevant text entries, missed recordings are possible. The data available through EMRALD are a voluntary sample of Ontarian physicians who all use one type of EMR system and practice under some type of primary care reform model of care, and therefore may not be entirely representative of all physicians in the province. For instance, patients captured by EMRALD are more likely to live in rural areas and to be cared for by younger physicians who have adopted EMR systems.[3] In summary, we have identified that our algorithm, utilizing administrative data, has good sensitivity, and excellent specificity, to accurately characterize pertussis immunization status in infants. However, the sensitivity and specificity declined with age.

References

1. Ducharme R, Benchimol EI, Deeks SL, Hawken S, Fergusson DA, Wilson K. Validation of Diagnostic Codes for Intussusception and Quantification of Childhood Intussusception Incidence in Ontario, Canada: A Population-Based Study. The Journal of Pediatrics **2013**; 163(4): 1073-9.e3.
2. Publically funded immunization schedule in Ontario (2011): Accessed May 30 2014. Available from: http://www.health.gov.on.ca/en/public/programs/immunization/docs/schedule.pdf.
3. Schwartz KL, Tu K, Wing L, et al. Validation of infant immunization billing codes in administrative data. Hum Vaccin Immunother **2015**; 11(7):1840-7.
4. Tu K, Mitiku TF, Ivers NM, et al. Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD). Am J Manag Care **2014**; 20(1): e15-21.
5. Ontario MD EMR Enhanced Use Program. [accessed 2014 December 11]; Available from: https://http://www.ontariomd.ca/portal/server.pt/community/emr_funding/emr_enhanced_use_program/.
6. Tu K, Mitiku T, Lee DS, Guo H, Tu JV. Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD). Can J Cardiol **2010**; 26(7): e225-8.
7. Tu K, Wang M, Young J, et al. Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMRALD as a reference standard. Can J Cardiol **2013**; 29(11): 1388-94.
8. Tu K, Wang M, Jaakkimainen RL, et al. Assessing the validity of using administrative data to identify patients with epilepsy. Epilepsia **2014**; 55(2): 335-43.

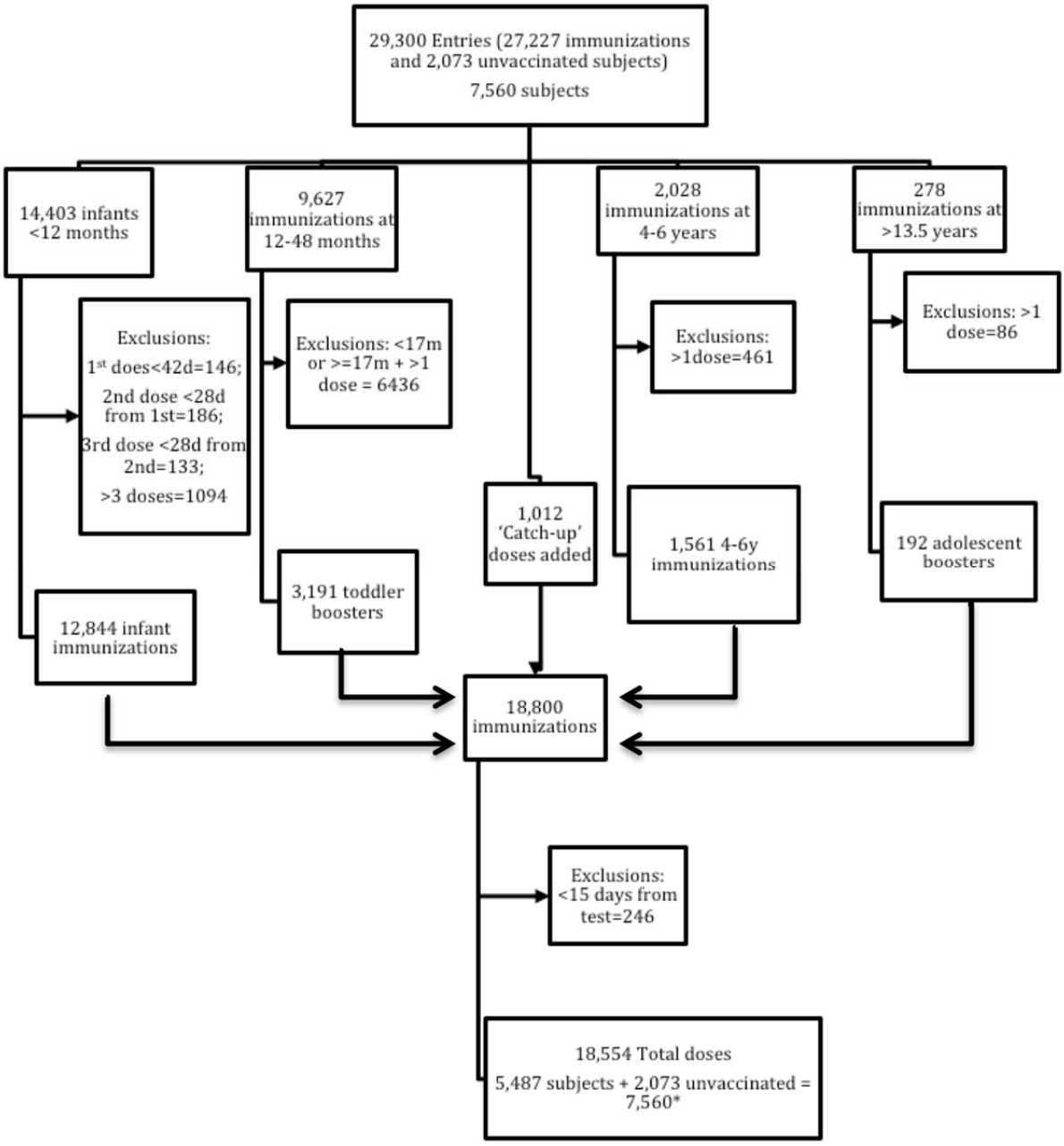

Figure e1: Flow diagram of applying the pertussis vaccine algorithm to the OHIP immunization database.
*****NB: includes infants 0-2 months

Table e1: Evaluating pertussis vaccine algorithm for Ontario Health Insurance Plan physician billing claims performance characteristics with 95% confidence intervals compared with electronic medical records as the reference standard.

	Cell Counts				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
	TP	FP	FN	TN				
UTD for pertussis at 3 months	1246	5	171	89	87.9 (86.2-89.6)	94.7 (90.1-99.2)	99.6 (99.3-99.9)	34.2 (28.5-40.0)
UTD for pertussis at 5 months	1097	9	255	180	83.0 (81.0-85.0)	95.2 (92.2-98.3)	99.2 (98.7-99.7)	44.4 (39.6-49.3)
UTD for pertussis at 7 months	953	10	255	293	78.9 (76.6-81.2)	96.7 (94.7-98.7)	99.0 (98.3-99.6)	53.5 (49.3-57.7)
UTD for pertussis at 19 months	598	15	310	588	65.9 (62.8-68.9)	97.5 (96.3-98.8)	97.6 (96.3-98.8)	65.5 (62.4-68.6)
UTD for pertussis at 7 years	739	38	551	183	57.3 (54.5-60.0)	82.8 (77.8-87.8)	95.1 (93.6-96.6)	24.9 (21.8-28.1)
Unvaccinated	6	3	3	1499	66.7 (35.9-97.5)	99.8 (99.6-100)	66.7 (35.9-97.5)	99.8 (99.6-100)

TP=True Positive (vaccination in OHIP database and EMRALD)
FP=False Positive (vaccination in OHIP database but not in EMRALD)
FN=False Negative (vaccination in EMRALD but not in OHIP database)
TN=True Negative (no vaccinatrion in OHIP database or EMRALD)
UTD=up-to-date; PPV=positive predictive value; NPV=negative predictive value; CI=confidence interval