Automating the Surveillance of Healthcare-Associated Infections

Maaike S.M. van Mourik

Automating the Surveillance of Healthcare-Associated Infections
PhD thesis, University of Utrecht, the Netherlands
ISBN: 978-90-5335-906-8
Lay-out: Nikki Vermeulen, Ridderprint BV
Cover lay-out: Elise van Mourik
Printing: Ridderprint, Ridderkerk, the Netherlands
© Maaike van Mourik, Utrecht, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. The copyright of articles that have been published or accepted for publication has been transferred to the respective journals.

The research presented in this thesis was in part supported by an Alexander Suermann Stipend provided by the University Medical Center Utrecht.

The printing of this thesis was kindly supported by the Julius Center for Health Sciences and Primary Care, SAS Nederland B.V., the Netherlands Society of Medical Microbiology (NVMM) and the Royal Netherlands Society for Microbiology (KNVM) and Chipsoft BV.

Automating the Surveillance of Healthcare-Associated Infections

Geautomatiseerde Surveillance van Zorginfecties
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 30 september 2014 des middags te 4.15 uur
door

Maaike Sybil Margareta van Mourik

geboren op 20 oktober 1985
te Zeist

Promotoren:	Prof.dr. M.J.M. Bonten Prof.dr. K.G.M. Moons
Copromotor:	Dr. A. Troelstra

Leden van de leescommissie:	Prof.dr. M.M.E. Schneider Prof.dr. C.J. Kalkman Prof.dr. J.A.J.W. Kluytmans Prof.dr. K.C.B. Roes Prof.dr. M.C. Vos
Paranimfen	Evelien T.M. Berends Nienke L. Plantinga

TABLE OF CONTENTS

1

General introduction

Maaike S.M. van Mourik

Healthcare-associated infections (HAI) are infections that patients acquire during their course of medical treatment for other conditions. As part of an increased awareness for patient safety, also propelled by the publication of the seminal report 'To Err is Human', the surveillance and prevention of HAI has received growing attention in recent years[1,2]. HAI can occur in any organ system and may be related to operative procedures or invasive devices such as central vascular catheters, urinary catheters or mechanical ventilation. These infections may sometimes be caused by antimicrobial-resistant micro-organisms, although this is not necessarily the case[3-5].

The worldwide prevalence of HAI has been estimated at 7.6% of patients admitted to hospitals in high-income countries, and incidence is notably higher in Intensive Care Units (ICUs)[6]. In a recent European prevalence survey, an episode of HAI was diagnosed in 6.0% of all hospitalized patients on any given day, translating to approximately 3.2 million HAI annually. In this study, respiratory tract infections were most common, followed by surgical site infections, urinary tract infections and bloodstream infections[7,8]. In the United States (U.S.), a recent prevalence survey targeting the full scope of HAI found that one in 25 patients admitted to acute care hospitals suffers from a HAI, with a calculated incidence of 730'000 infections yearly[9,10].

Acquisition of HAI leads to increased antimicrobial exposure, length of hospital stay and risk of mortality in affected patients[6,11,12]. Aside from causing incremental suffering for individual patients, HAI pose a financial burden on healthcare systems. In the U.S., the direct medical costs for the most common HAI were estimated at 9.8 billion dollars annually[13], somewhat lower than previous studies that targeted a broader scope of HAI (28 to 45 billion dollars)[14]. Accounting for societal costs and loss of lives further increases this estimate, and this is not different in other high-income countries[15,16].

HAI surveillance

The landmark Study on the Efficacy of Nosocomial Infection Control performed in the 1980s underscored the importance of HAI surveillance and the feedback of infection rates to stakeholders as a cornerstone of prevention programs[17]. Numerous studies from various countries have since shown similar results[18-20]. Surveillance entails the "systematic collection, analysis, interpretation and dissemination of data regarding a health-event for use in public health action to reduce morbidity and mortality and to improve health"[21]. Many countries have organized HAI surveillance networks, such as the *National Healthcare Safety Network* (NHSN) in the U.S., the *PREventie van ZIEkenhuisinfecties door Surveillance* (PREZIES) initiative in the Netherlands and the German *Krankenhaus Infektions Surveillance System* (KISS). These networks aim to provide standardized

surveillance methodology, collect and aggregate data and provide reference information to facilitate hospitals and policy makers in the interpretation of HAI rates, support the implementation of prevention strategies and identify opportunities for improvement[6,7,22-24].

Importantly, data collected on HAI incidence is increasingly used not only as a means to drive quality improvement by voluntary and confidential feedback to individual hospitals, but also as a quality indicator to evaluate healthcare providers. Participation in surveillance networks has become mandatory or associated with payment incentives in a number of countries, and public disclosure of HAI rates is increasingly advocated[25-27]; for example, HAI are among the measures published on the *Hospital Compare* website in the United States[28]. In addition, pay-for-performance programs have been developed that impose financial penalties on the occurrence of HAI[29]. This increasing use of HAI surveillance systems to evaluate quality is in keeping with patient safety initiatives targeting other types of adverse events and patient outcomes[30,31]. For example, registries evaluating patient outcomes after surgical procedures or critical care are increasingly being adopted worldwide, with healthcare providers being rewarded for participation and/or observed performance[32-34]. Although heavily debated, public disclosure of patient outcomes such as hospital standardized mortality ratios is becoming the standard[27,28,35-37].

Methods of HAI surveillance

The surveillance of HAI needed for quality improvement programs requires reliable ascertainment of whether a patient developed an infection during the course of medical treatment. Traditionally, patients' medical records are reviewed manually – prospectively or retrospectively – and standardized case-definitions are applied to determine HAI presence[4,6,38]. This manual chart review gathers information from a diversity of sources, including clinical notes, microbiology culture results, antibiotic dispensing records, surgical information, radiology results and, for some HAI, post-discharge follow-up information[6,39,40].

To be meaningful to clinicians, patients, policy makers and other stakeholders, surveillance programs must reliably measure the presence of HAI. Key components of such a reliable measurement are accurate and timely identification of clinically relevant infections in a manner that is consistent across reviewers and time[41]. In addition, if surveillance data are to be used for interfacility comparison or reimbursement programs, transportability and consistency across different hospitals must be ensured along with adequate methods of correcting for underlying differences between patient populations[36,42,43].

Although traditional surveillance of HAI by manual methods is considered the reference method, it is hampered by several limitations. Not only is manual review of patients records time-consuming and therefore of limited capacity, many questions have been raised regarding the reliability of surveillance efforts[1,41,44-46]. Threats to reliability of HAI detection can originate in a number of ways. Surveillance requires collection of information from a myriad of data sources to allow for application of case-definitions; this data collection process is effort-dependent and suffers from the premise that 'the more you look, the more you find'[47,48]. Moreover, (healthcare-associated) infections can have many clinical presentations and application of case-definitions is not straightforward resulting in suboptimal interrater reliability, even when providing assessors with prepared case-vignettes[49-51]. Finally, there may be systematic differences between hospitals or across time in implementation of surveillance within the process of care, either driven by variation in interpretation of definitions or other incentives such as 'a wish to do good', attempts to generate more meaningful rates or in reaction to financial penalties[27,52-54].

Routine care data

These concerns with traditional surveillance methods together with increasing adoption of electronic health records (EHRs) and sophisticated methods of data management have motivated the development of surveillance systems using electronic routine care data since the 1980s[55-58]. Routine care data include all patient information collected and stored during the process of clinical care, comprising both data collected for administrative purposes (e.g., diagnosis codes used for billing) and clinical patient data such as demographic information, clinical characteristics, diagnostic testing and pharmacy records[59,60]. Clinical data warehouses link these multiple sources of information, and provide access to routine care data, either in structured, well-defined fields or as unstructured free-text[60-63].

Current methods of (semi-)automated surveillance

Routine care data have been utilized in several ways to support HAI surveillance. Most automated surveillance systems aim to identify probable cases of HAI, for example by classifying patients with an indicator of infection as high-risk and limiting manual review to these records. Commonly included indicators of infection are microbiology results, admission and discharge records, antibiotic use, procedure codes and administrative data and some systems incorporate data from the clinical laboratory or apply natural language processing of free-text[56,59,64-69]. Other approaches use multivariable regression models or employ concepts such as fuzzy logic to discern possible episodes of HAI[56,70,71].

As has previously been underlined by others, automated surveillance systems can have different intended applications. They may either be used as a means to support traditional surveillance methods within individual hospitals by consistently identifying high-risk patients requiring manual chart review or serve as a fully automated system aiming to achieve a definite diagnosis of HAI, possibly using identical methods in multiple hospitals with the purpose of outcome comparisons[72].

A recent survey indicated that approximately one third of U.S. hospitals employ some form of automated surveillance to support their efforts, and adoption of such systems has been associated with successful infection prevention programs[44,73,74]. Several studies have shown advantages of automated HAI surveillance systems, not only with regards to time-savings and hence increased capacity[75,76] but also the capability to increase the reliability of the surveillance process, for example by removing the effort-dependent nature of data collection and ensuring more consistent case-identification[45,77]. Of the systems developed to support in-hospital surveillance, sensitivities over 80% have been reported together with 60 – 80% workload reductions and comparison of hospitals based on automated surveillance systems that use clinical routine care data may be superior to manual surveillance[71,76-79]. Administrative data, in particular, are also being used to develop performance indicators for the direct comparison across hospitals[80,81].

Thesis aim and outline

Although progress has been made towards automation of HAI surveillance, unanswered questions do remain. Many of the existing systems rely on a narrow set of information sources to identify high-risk patients, and the surveillance approach with the most favorable sensitivity-to-efficiency tradeoff remains to be identified, either by including different types of routine care data or by using alternative modeling methods. Also the robustness of automated surveillance systems to differences in clinical practice or minor variations in data specification requires further assessment. Hence, the aim of this thesis is to further explore the possibilities and limitations of automated HAI surveillance using routine care data.

The first part of this thesis focuses on harnessing clinical routine care data for HAI surveillance by developing an automated surveillance strategy to support within-hospital surveillance for a prototype infection. **Chapter 2** presents the development of a surveillance approach for drain-related meningitis, and the results of a temporal validation and update of this approach, are presented in **chapter 3**. **Chapter 4** provides a review of existing automated surveillance systems and places the results of chapters 2 and 3 in context by comparing several approaches to automated HAI surveillance as well

as discussing some remaining challenges. **Chapter 5** presents the results of a multicenter validation of the developed surveillance method for drain-related meningitis, again contrasted to an alternative approach to automated surveillance and assessing both the accuracy of the method and the generalizability across different hospitals.

Concerns regarding the reliability of traditional manual surveillance have also driven a movement towards the development of more objective HAI case-definitions that are amenable to electronic implementation using clinical routine care data[82]. **Chapter 6** presents a validation study of novel surveillance definitions for complications of mechanical ventilation and, in addition, focuses on the reliability of electronic implementation. **Chapter 7** presents the use of routine care data for a further component necessary for interpretation of HAI rates, namely correcting for differences in severity of underlying disease, focusing specifically on mechanically ventilated patients.

In the second part of this thesis, the focus shifts to using administrative routine care data for the purpose of HAI surveillance. In **chapter 8** the accuracy of discharge codes and billing information for detection of drain-related meningitis is assessed and compared with automated surveillance strategies using clinical data sources. **Chapter 9** presents a broader perspective on the use of administrative data for HAI surveillance; this systematic review targets the full scope of HAI and summarizes evidence regarding the possible applications and pitfalls of surveillance based on discharge coding.

A synthesis of the results and general discussion on surveillance of HAI using routine care data is provided in **Chapter 10**.

REFERENCE LIST

1. Burke JP. Infection control - a problem for patient safety. *N Engl J Med.* **2003**;348:651-656.
2. Institute of Medicine. *To Err is Human: Building a Safer Health System*. **2000**.
3. Centers for Disease Control and Prevention. Healthcare-associated infections (HAIs). **2013**. http://www.cdc.gov/HAI/burden.html. Last accessed: Mar 11 2014.
4. Centers for Disease Control and Prevention. CDC/NHSN Surveillance definitions for specific types of infection. **2014**. http://www.cdc.gov/nhsn/pdfs/pscmanual/17pscnosinfdef_current.pdf. Last accessed: Mar 11 2014.
5. World Health Organization. Prevention of hospital-acquired infections: a practical guide. **2002** Geneva
6. World Health Organization. Report on the Burden of Endemic Health Care-Associated Infection Worldwide. **2011**. http://apps.who.int/iris/bitstream/10665/80135/1/9789241501507_eng.pdf. Last accessed: Mar 14 2014.
7. European Centers for Disease Control. Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals 2011 - 2012. **2013**
8. PREZIES. Referentiecijfers maart 2007 t/m maart 2012: Prevalentieonderzoek. **2012**
9. Magill SS, Edwards JR, Bamberg W et al. Multistate point-prevalence survey of health care-associated infections. *N Engl J Med.* **2014**;370:1198-1208.
10. Klevens RM, Edwards JR, Richards CL, Jr. et al. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public Health Rep.* **2007**;122:160-166.
11. Lambert ML, Suetens C, Savey A et al. Clinical outcomes of health-care-associated infections and antimicrobial resistance in patients admitted to European intensive-care units: A cohort study. *Lancet Infect Dis.* **2011**;11:30-38.
12. Melsen WG, Rovers MM, Groenwold RH et al. Attributable mortality of ventilator-associated pneumonia: a meta-analysis of individual patient data from randomised prevention studies. *Lancet Infect Dis.* **2013**;13:665-671.
13. Zimlichman E, Henderson D, Tamir O et al. Health Care-Associated Infections: A Meta-analysis of Costs and Financial Impact on the US Health Care System. *JAMA Intern Med.* **2013**;173:2039-2046.
14. Scott, R. D. The direct medical cost of healthcare-associated infections in U.S. hospitals and the benefits of prevention. **2009**. http://www.cdc.gov/HAI/pdfs/hai/Scott_CostPaper.pdf. Last accessed: Jan 23 2012.
15. Roberts RR, Scott RD, Hota B et al. Costs attributable to healthcare-acquired infection in hospitalized adults and a comparison of economic methods. *Med Care.* **2010**;48:1026-1035.
16. Marchetti A, Rossiter R. Economic burden of healthcare-associated infection in US acute care hospitals: societal perspective. *J Med Econ.* **2013**;16:1399-1404.
17. Haley RW, Culver DH, White JW et al. The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals. *Am J Epidemiol.* **1985**;121:182-205.
18. Geubbels EL, Nagelkerke NJ, Mintjes-De Groot AJ, Vandenbroucke-Grauls CM, Grobbee DE, De Boer AS. Reduced risk of surgical site infections through surveillance in a network. *Int J Qual Health Care.* **2006**;18:127-133.
19. Brandt C, Sohr D, Behnke M, Daschner F, Ruden H, Gastmeier P. Reduction of surgical site infection rates associated with active surveillance. *Infect Control Hosp Epidemiol.* **2006**;27:1347-1351.
20. Gaynes R, Richards C, Edwards J et al. Feeding back surveillance data to prevent hospital-acquired infections. *Emerg Infect Dis.* **2001**;7:295-298.
21. U.S.Department of Health and Human Services. Updated Guidelines for Evaluating Public Health Surveillance Systems. *MMWR Morb Mortal Wkly Rep.* **2001**;50:1-35.
22. Centers for Disease Control and Prevention. National Healthcare Safety Network (NHSN). **2014**. http://www.cdc.gov/nhsn/. Last accessed: Mar 15 2014.
23. Rijksinstituut voor Volksgezondheid en Mileu. PREZIES: PREventie van ZIEkenhuisinfections door Surveillance. **2013**. http://www.rivm.nl/Onderwerpen/P/PREZIES. Last accessed: Mar 15 2014.

24. Nationales Referenzzentrum für Surveillance von nosokomialen Infektionen. KISS Project description. **2013**. http://www.nrz-hygiene.de/en/surveillance/hospital-infection-surveillance-system/. Last accessed: Mar 15 2014.
25. Haustein T, Gastmeier P, Holmes A et al. Use of benchmarking and public reporting for infection control in four high-income countries. *Lancet Infect Dis.* **2011**;11:471-481.
26. Department of Health and Human Services, Centers for Medicare and Medicaid Services. FY 2014 IPPS Final Rule Medicare Program: Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status. *Federal Register.* **2013**;78:50495-51040.
27. Talbot TR, Bratzler DW, Carrico RM et al. Public reporting of health care-associated surveillance data: recommendations from the healthcare infection control practices advisory committee. *Ann Intern Med.* **2013**;159:631-635.
28. Centers for Medicare and Medicaid Services. Medicare Hospital Compare Quality of Care. **2014**. http://www.medicare.gov/hospitalcompare/search.html. Last accessed: Mar 20 2014.
29. Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med.* **2007**;357:1573-1575.
30. Agency for Healthcare Quality and Research. National Quality Forum (NQF) Endorsed Individual and Composite Measures. **2013**. http://www.qualityindicators.ahrq.gov/Downloads/Modules/V45/Module_NQF_Endorsement_V4.5.pdf. Last accessed: Mar 20 2014.
31. VMS Veiligheidsprogramma. VMS Thema's. **2013**. http://www.vmszorg.nl/Themas. Last accessed: Mar 15 2014.
32. Dutch Institute for Clinical Auditing. DICA Rapportages 2012. **2013**
33. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg.* **2009**;250:363-376.
34. Stichting NICE. Nationale Intensive Care Evaluatie. **2013**. http://www.stichting-nice.nl/index.jsp. Last accessed: Mar 20 2014.
35. Health Council of the Netherlands (Gezondheidsraad). Publieke indicatoren voor de kwaliteit van curatieve zorg: De stand van de discussie. **2013**
36. Panzer RJ, Gitomer RS, Greene WH, Webster PR, Landry KR, Riccobono CA. Increasing demands for quality measurement. *JAMA*. **2013**;310:1971-1980.
37. Nederlandse Zorgautoriteit. Regeling verplichte publicatie sterftecijfers instellingen voor medisch specialistische zorg. **2014**
38. PREZIES. Bijlage 2: Definities ziekenhuisinfecties. **2012**. http://www.prezies.nl/zkh/prev/definities.html. Last accessed: May 5 2014.
39. Centers for Disease Control and Prevention and Center. Outline for Healthcare-Associated Infection Surveillance. **2006**. http://www.cdc.gov/nhsn/pdfs/outlineforHAIsurveillance.pdf. Last accessed: Mar 25 2014.
40. Mannien J, Wille JC, Snoeren RL, van den HS. Impact of postdischarge surveillance on surgical site infection rates for several surgical procedures: results from the nosocomial surveillance network in The Netherlands. *Infect Control Hosp Epidemiol.* **2006**;27:809-816.
41. Tokars JI, Richards C, Andrus M et al. The changing face of surveillance for health care-associated infections. *Clin Infect Dis.* **2004**;39:1347-1352.
42. Harris AD, McGregor JC. The importance of case-mix adjustment for infection rates and the need for more research. *Infect Control Hosp Epidemiol.* **2008**;29:693-694.
43. Farmer SA, Black B, Bonow RO. Tension between quality measurement, public quality reporting, and pay for performance. *JAMA*. **2013**;309:349-350.
44. Stone PW, Pogorzelska-Maziarz M, Herzig CT et al. State of infection prevention in US hospitals enrolled in the National Health and Safety Network. *Am J Infect Control.* **2014**;42:94-99.
45. Trick WE. Decision making during healthcare-associated infection surveillance: a rationale for automation. *Clin Infect Dis.* **2013**;57:434-440.

46. Emori TG, Edwards JR, Culver DH et al. Accuracy of reporting nosocomial infections in intensive-care-unit patients to the National Nosocomial Infections Surveillance System: a pilot study. *Infect Control Hosp Epidemiol.* **1998**;19:308-316.
47. Niedner MF. The harder you look, the more you find: Catheter-associated bloodstream infection surveillance variability. *Am J Infect Control.* **2010**;38:585-595.
48. Haut ER, Pronovost PJ. Surveillance bias in outcomes reporting. *JAMA.* **2011**;305:2462-2463.
49. Stevens JP, Kachniarz B, Wright SB et al. When Policy Gets It Right: Variability in U.S. Hospitals' Diagnosis of Ventilator-Associated Pneumonia. *Crit Care Med.* **2013**;42:497-503.
50. Keller SC, Linkin DR, Fishman NO, Lautenbach E. Variations in identification of healthcare-associated infections. *Infect Control Hosp Epidemiol.* **2013**;34:678-686.
51. Mayer J, Greene T, Howell J et al. Agreement in Classifying Bloodstream Infections Among Multiple Reviewers Conducting Surveillance. *Clin Infect Dis.* **2012**;55:364-370.
52. Klompas M. Eight initiatives that misleadingly lower ventilator-associated pneumonia rates. *Am J Infect Control.* **2012**;40:408-410.
53. Dixon-Woods M, Leslie M, Bion J, Tarrant C. What counts? An ethnographic study of infection data reported to a patient safety program. *Milbank Q.* **2012**;90:548-591.
54. Tanner J, Padley W, Kiernan M, Leaper D, Norrie P, Baggott R. A benchmark too far: findings from a national survey of surgical site infection surveillance. *J Hosp Infect.* **2013**;83:87-91.
55. Evans RS, Larsen RA, Burke JP et al. Computer surveillance of hospital-acquired infections and antibiotic use. *JAMA.* **1986**;256:1007-1011.
56. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. *Clin Infect Dis.* **2009**;48:1268-1275.
57. Atreja A, Gordon SM, Pollock DA, Olmsted RN, Brennan PJ. Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. *Am J Infect Control.* **2008**;36:S37-S46.
58. Jha AK, DesRoches CM, Campbell EG et al. Use of electronic health records in U.S. hospitals. *N Engl J Med.* **2009**;360:1628-1638.
59. De Bruin JS, Seeling W, Schuh C. Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review. *J Am Med Inform Assoc.* **2014**;In Press.
60. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med.* **2007**;45:13-19.
61. Trick WE. Building a data warehouse for infection control. *Am J Infect Control.* **2008**;36:S75-81.
62. Garcia AL, Aylin P, Tian J et al. Data linkage between existing healthcare databases to support hospital epidemiology. *J Hosp Infect.* **2011**;79:231-235.
63. Murff HJ, FitzHenry F, Matheny ME et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA.* **2011**;306:848-855.
64. Freeman R, Moore LS, Garcia AL, Charlett A, Holmes A. Advances in electronic surveillance for healthcare-associated infections in the 21st Century: a systematic review. *J Hosp Infect.* **2013**;84:106-119.
65. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. *Clin Infect Dis.* **2009**;49:949-955.
66. Leal J, Laupland KB. Validity of electronic surveillance systems: a systematic review. *J Hosp Infect.* **2008**;69:220-229.
67. Haas JP, Mendonca EA, Ross B, Friedman C, Larson E. Use of computerized surveillance to detect nosocomial pneumonia in neonatal intensive care unit patients. *Am J Infect Control.* **2005**;33:439-443.
68. Leth RA, Moller JK. Surveillance of hospital-acquired infections based on electronic hospital registries. *J Hosp Infect.* **2006**;62:71-79.
69. King C, Aylin P, Moore LS, Pavlu J, Holmes A. Syndromic surveillance of surgical site infections - A case study in coronary artery bypass graft patients. *J Infect.* **2014**;68:23-31.
70. van Mourik MS, Troelstra A, van Solinge WW, Moons KG, Bonten MJ. Automated surveillance for healthcare-associated infections: opportunities for improvement. *Clin Infect Dis.* **2013**;57:85-93.

71. De Bruin JS, Adlassnig KP, Blacky A, Mandl H, Fehre K, Koller W. Effectiveness of an automated surveillance system for intensive care unit-acquired infections. *J Am Med Inform Assoc.* **2013**;20:369-372.
72. Woeltje KF. Moving into the future: electronic surveillance for healthcare-associated infections. *J Hosp Infect.* **2013**;84:103-105.
73. Halpin H, Shortell SM, Milstein A, Vanneman M. Hospital adoption of automated surveillance technology and the implementation of infection prevention and control programs. *Am J Infect Control.* **2011**;39:270-276.
74. Palumbo AJ, Loveless PA, Moll ME, Ostroff S. Evaluation of healthcare-associated infection surveillance in pennsylvania hospitals. *Infect Control Hosp Epidemiol.* **2012**;33:105-111.
75. Chalfine A, Cauet D, Lin WC et al. Highly sensitive and efficient computer-assisted system for routine surveillance for surgical site infection. *Infect Control Hosp Epidemiol.* **2006**;27:794-801.
76. Knepper BC, Young H, Jenkins TC, Price CS. Time-saving impact of an algorithm to identify potential surgical site infections. *Infect Control Hosp Epidemiol.* **2013**;34:1094-1098.
77. Lin MY, Hota B, Khan YM et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. *JAMA.* **2010**;304:2035-2041.
78. Hayashi Y, Morisawa K, Klompas M et al. Toward Improved Surveillance: The Impact of Ventilator-Associated Complications on Length of Stay and Antibiotic Use in Patients in Intensive Care Units. *Clin Infect Dis.* **2013**;56:471-477.
79. Bolon MK, Hooper D, Stevenson KB et al. Improved surveillance for surgical site infections after orthopedic implantation procedures: extending applications for automated data. *Clin Infect Dis.* **2009**;48:1223-1229.
80. Huang SS, Livingston JM, Rawson NS, Schmaltz S, Platt R. Developing algorithms for healthcare insurers to systematically monitor surgical site infection rates. *BMC Med Res Methodol.* **2007**;7:20.
81. Agency for Healthcare Research and Quality. Patient Safety Indicators Overview. **2014**. http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx. Last accessed: Mar 18 2014.
82. Magill SS, Klompas M, Balk R et al. Developing a new, national approach to surveillance for ventilator-associated events*. *Crit Care Med.* **2013**;41:2467-2475.

Part 1

Harnessing clinical patient data

2

Automated detection of external ventricular and lumbar drain-related meningitis using laboratory and microbiology results and medication data

Maaike S.M. van Mourik[a], Rolf H.H. Groenwold[b], Jan Willem Berkelbach van der Sprenkel[c], Wouter W. van Solinge[d], Annet Troelstra[a], Marc J.M. Bonten[a, b]

[a] *Department of Medical Microbiology*
[b] *Julius Center for Health Sciences and Primary Care*
[c] *Department of Neurosurgery*
[d] *Department of Clinical Chemistry and Hematology*
All at University Medical Center Utrecht, Utrecht, The Netherlands

PLoS ONE 2011 6(8):e22846.

ABSTRACT

Objective – Monitoring of healthcare-associated infection rates is important for infection control and hospital benchmarking. However, manual surveillance is time-consuming and susceptible to error. The aim was, therefore, to develop a prediction model to retrospectively detect drain-related meningitis (DRM), a frequently occurring nosocomial infection, using routinely collected data from a clinical data warehouse.

Methods – As part of the hospital infection control program, all patients receiving an external ventricular (EVD) or lumbar drain (ELD) (2004 to 2009; n=742) had been evaluated for the development of DRM through chart review and standardized diagnostic criteria by infection control staff; this was the reference standard. Children, patients dying <24 hours after drain insertion or with <1 day follow-up and patients with infection at the time of insertion or multiple simultaneous drains were excluded. Logistic regression was used to develop a model predicting the occurrence of DRM. Missing data were imputed using multiple imputation. Bootstrapping was applied to increase generalizability.

Results – 537 patients remained after application of exclusion criteria, of which 82 developed DRM (13.5/1000 days at risk). The automated model to detect DRM included the number of drains placed, drain type, blood leukocyte count, C-reactive protein, cerebrospinal fluid leukocyte count and culture result, number of antibiotics started during admission, and empiric antibiotic therapy. Discriminatory power of this model was excellent (area under the ROC curve 0.97). The model achieved 98.8% sensitivity (95% CI 88.0% to 99.9%) and specificity of 87.9% (84.6% to 90.8%). Positive and negative predictive values were 56.9% (50.8% to 67.9%) and 99.9% (98.6% to 99.9%), respectively. Predicted yearly infection rates concurred with observed infection rates.

Conclusion – A prediction model based on multi-source data stored in a clinical data warehouse could accurately quantify rates of DRM. Automated detection using this statistical approach is feasible and could be applied to other nosocomial infections.

INTRODUCTION

Healthcare-associated infections (HAI) pose a great burden on current medical care and are increasingly viewed as preventable complications. HAI refers to the entire scope of infections associated with medical care and also includes nosocomial infections. The European burden of HAI has been estimated at 4.5 million infections contributing to 148,000 deaths[1]. Hospitals are encouraged to report HAI rates through surveillance organizations such as the National Healthcare Safety Network (NHSN) from the Centers for Disease Control and Prevention (CDC) in the United States[2] and the PREZIES network in the Netherlands[1,3]. Efficient registration and feedback of infection rates to healthcare workers are considered essential elements to reduce infection rates and surveillance of infection rates is increasingly demanded by policy makers and the public[4]. However, manual registration of HAI rates is time-consuming and susceptible to error due to subjective interpretation of definitions and manual data handling[5,6].

Therefore, there is an urgent need for more efficient and reliable surveillance methods. Automated classification algorithms using data stored in electronic medical records (i.e. clinical data warehouses[7]) have been developed both at the hospital-wide and procedure-specific level with varying success, in particular for surgical site infections and (catheter-related) bloodstream infections[8-18]. Most models use a classification approach based on the presence of (one or more) indicators of infection, such as positive microbiology results, antibiotic use, and discharge coding. Another, less used, method is development of a multivariable model with associated cut-off values to classify patients[11,13]. Such (automated) models are of relatively low cost, time-saving, and facilitate standardized interpretation of infection criteria[10]. However, when case-finding is based on microbiological cultures, such classification algorithms have low sensitivity for culture-negative infections and specificity of these algorithms decreases when extending case-finding criteria. Furthermore, suboptimal positive predictive values still make manual confirmation of infection necessary. Given these potential drawbacks, most healthcare centers still use manual chart review as their primary method of surveillance.

The aim of this study was to develop a prediction model for HAI using data routinely stored in a clinical data warehouse (i.e. the Utrecht patient-oriented database, UPOD[19]) to retrospectively identify the presence of infection. In order to increase sensitivity for culture-negative infections while maintaining specificity and enable surveillance without the need for manual confirmation, an alternative approach to the classification algorithm was sought and data sources were extended to include not only microbiology results and antibiotic use but also results of clinical chemistry analysis. In clinical practice, such a model could eliminate or significantly reduce the workload of manual chart review and

increase resources available for development and implementation of infection control measures. Drain-related meningitis (DRM) was selected as an example to investigate this general approach to automated evaluation of infection rates. This nosocomial infection, related to external cerebrospinal fluid (CSF) drainage through external ventricular (EVD) or lumbar drains (ELD), is sometimes also termed ventriculitis or meningoventriculitis and is one of the procedure-specific infections that has since 2004 been monitored by the department of hospital hygiene and infection control through labor-intensive manual chart review. The developed model achieved good discriminatory power at the level of the individual patient and group-level estimates of infection rates could be generated without any manual confirmation.

METHODS

Ethics statement

The use of anonymous data through the UPOD has been exempted from review by the Institutional Review Board of the University Medical Center Utrecht as described previously[19].

Study design and outcome measure

Data collected as part of the hospital hygiene surveillance program were used to develop the prediction model for DRM. Results of routinely performed incidence surveys were considered as reference standard. Two infection control professionals assessed each patient for the development of DRM by chart review using modified NHSN/CDC criteria for healthcare-associated meningitis (**figure 1**)[20,21]; in case of disagreement adjudication was performed through review. A surveillance episode was defined to start the day of drain placement up to seven days after drain removal of the last drain or up to discharge, whichever occurred first.

Study population

All patients registered by the surveillance program to have received an external cerebrospinal fluid drain at the University Medical Centre Utrecht, a 1042-bed tertiary healthcare centre, were included in this study. Registration comprises all patients who received an EVD between January 1st 2004 and December 31st 2009 (with the exception of May-July 2004), and all patients receiving and ELD in 2004 to 2006. From January 2007 to December 2009 surveillance for ELD was only performed in patients who received the drain in operating theatres. Several infection control measures were implemented during the study period as described previously[20]. All EVDs are placed in operating theatres or,

NHSN/CDC definition of healthcare-associated meningitis for patients > 1 year of age
At least 1 of the following criteria:

(1) Patient has organism cultured from cerebrospinal fluid (CSF).

(2) Patient has at least one of the following symptoms with no other recognized cause: fever (>38°C), headache, stiff neck, meningeal signs, cranial nerve signs, or irritability.
And at least 1 of the following:
a. Increased white cells, elevated protein and/or decreased glucose in CSF.
b. Organisms seen on Gram stain of CSF.
c. Organisms cultured from blood.
d. Positive antigen test from blood, CSF or urine.
e. Diagnostic single antibody titer or 4-fold increase for pathogen.
And, if diagnosis is made antemortem, physician instituted appropriate antimicrobial therapy.

Modifications:
- Coagulase-negative staphylococci are classified as contamination if further clinical or biochemical abnormalities are absent.
- Blood-contaminated CSF is interpreted as positive if the leukocyte to erythrocyte ratio exceeds 1:100.

Figure 1 | NHSN/CDC definition of healthcare-associated meningitis for patients > 1 year of age[20,21].

sometimes, under sterile conditions in the intensive care unit (ICU) by a neurosurgeon or trained resident. ELDs are either inserted in the operating theatre or in sterile conditions on the neurology ward. EVDs are tunneled five centimeters under the skin. All patients receive perioperative antibiotic prophylaxis. Drains are not exchanged on a prophylactic basis and CSF samples are collected for culture and biochemical analysis only when infection is clinically suspected; at this time empiric antibiotic therapy is initiated according to local protocol.

Throughout the study period, 742 patients received one or more drains. The following exclusion criteria were applied: death within one day after drain placement (40 patients), age under 18 years (n = 110), pre-existing central nervous system infection (n = 29), more than one simultaneous drain (n = 8), drain placement in a different centre (n = 4), admission duration less than one day (n = 3), second admission more than 30 days after discharge from a first drainage episode within the study (n = 7), and admission to the military hospital (n = 2). After application of exclusion criteria, 537 patients were available for analysis.

Data collection

The department of hospital hygiene provided outcome data along with the drain characteristics for each patient (drain type, duration, indication for placement). Prediction data was obtained through the Utrecht Patient Oriented Database (UPOD), a clinical data warehouse launched in 2004 for research purposes. The UPOD links results from laboratory analysis (clinical chemistry and hematology), microbiological

cultures, and pharmacy dispensing data to information from the hospital admission and discharge system at the patient level[19]. Medication prescription data were obtained both from the UPOD as well as directly from the intensive care unit. Data were adapted to a standardized format and checked for inconsistencies. When necessary, original data sources were cross-referenced to exclude errors.

Predictor selection

Predictors were selected both on theoretical grounds and to best match the modified NHSN/CDC criteria. Duration of drainage, drain manipulation, subarachnoid hemorrhage, cerebrospinal fluid leakage, and other concomitant infections have been described as possible risk factors for meningitis previously[22]. Besides microbiological analysis of CSF and drains (after removal), biochemical markers of meningitis such as CSF leukocyte count, neutrophil count, glucose level, total protein level, and CSF/blood glucose ratio have been applied to the neurosurgical population with moderate success[23,24] and were therefore considered as predictors. Unfortunately, Gram-stain results were not yet available and the UPOD does not contain information on drain manipulation and the occurrence of concomitant infections. If patients received both an EVD and an ELD, the EVD took priority in determining drain type. Culture results have been corrected for contamination by categorizing cultures growing coagulase-negative staphylococci as negative if no antibiotic therapy was initiated one day prior through three days after culture. Empiric antibiotic therapy was defined as the simultaneous use of vancomycine and ceftazidime (started four or more days after admission) or ceftriaxone and flucloxacillin (initiated within four days of admission) according to local protocol. The number of systemic antibiotics started throughout the surveillance episode was included as a surrogate marker for the presence of other concomitant infections.

Statistical analysis

Since the objective was to predict whether a patient had developed DRM during hospital stay, the value that was most indicative of infection measured throughout each patient's surveillance episode was taken for each predictor. Missing data were imputed using multiple imputation (ten imputations). For C-reactive protein (CRP), squared and cubic terms were included in the prediction model along with the linear term. The number of leukocytes in CSF was log-transformed prior to analyses.

Variables were selected for multivariate analysis based on theoretical considerations (as previously described) and results of univariate analysis ($p < 0.05$ in the mean dataset). Using logistic regression analysis, a prediction model was then developed by means

of manual backward selection ($p < 0.05$). Although it is recommended to use higher p-values for selection of predictors in prediction research[25], this more stringent criterion was used due to the limited number of events (higher p-values would have resulted in too few events per predictor). Regression coefficients and standard errors were determined on each imputation set and pooled using Rubin's rule[26]. Subsequently, bootstrapping (100 samples per imputation set) was applied to correct for optimism.

Discrimination and calibration were determined for the final model. Discrimination refers to the ability to distinguish between patients with and without DRM; this was assessed by the area under the ROC curve. Calibration refers to the concordance between the predicted and observed probabilities of infection, which was assessed using a calibration plot. For clinical application, cut-off values for a predicted probability associated with high sensitivity and acceptable specificity were determined and associated sensitivity, specificity and predictive values were reported. Confidence intervals were determined using exact binomial methods. Finally, the summed predicted probabilities were used to investigate infection rates at the group level. All analyses were done using SPSS® 17 (SPSS Inc, Chicago IL) and R version 2.11.1 (www.r-project.org).

RESULTS

A total of 691 drains were placed in 537 patients. DRM occurred in 82 patients (15.3%), or 13.5 infections per 1000 drainage days at risk. The most common causative micro-organisms were coagulase-negative staphylococci (33.8%), followed by *Staphylococcus aureus* (14.6%) and enterobacteriacaea (13.4%). Seventeen infections were culture negative (20.7%). Baseline characteristics are described in **table 1**. Median age of the included patients was 58.5 years, half (n = 263) received a CSF drainage system to treat secondary hydrocephalus following subarachnoid hemorrhage, intraventricular

Table 1 | Baseline characteristics of the patient population after multiple imputation of missing values and univariate association between variables and the risk of drain-related meningitis.

Median (IQR) or n (%)	Overall n = 537	No DRM n = 455	DRM n = 82	p-value[a]
Demographics				
Age (years)	58.5 (47.2-69.6)	59.3 (46.8-69.4)	56.0 (47.5-65.6)	0.49
Sex (% female)	290 (54.0)	247 (54.3)	43 (52.4)	0.78
In-hospital death (%)	90 (16.8)	79 (17.4)	11 (13.4)	0.38
Duration of admission (days)	21.0 (12.0-37.5)	19.0 (11.0-30.0)	40.0 (28.5-59.3)	<0.001
Admission on ICU (%)	312 (58.1)	253 (55.6)	59 (72.0)	0.006
Duration of ICU stay	2 (0.0-7.0)	2 (0.0-5.0)	4.5 (0.0-12.3)	<0.001

Table 1 | *Continued*

Median (IQR) or n (%)	Overall n = 537	No DRM n = 455	DRM n = 82	p-value[a]
Indication for first drain (%)				<0.001
- SAH/IVH	249 (46.4)	205 (42.0)	58 (70.7)	
- Infarction	14 (2.6)	14 (3.1)	0 (0)	
- CSF leakage	85 (15.8)	77 (16.9)	8 (9.8)	
- Perioperative	86 (16.0)	84 (18.5)	2 (2.4)	
- Trauma	14 (2.6)	11 (2.4)	3 (3.7)	
- Tumor	37 (6.9)	30 (6.6)	7 (8.5)	
- Other	52 (9.7)	48 (10.5)	4 (4.9)	
Drain characteristics				
Drain type (% EVD)	337 (62.8)	266 (58.5)	71 (86.6)	<0.001
Total drain duration (days)	9.0 (6.0-17.0)	8.0 (5.0-13.0)	20.0 (15.0-29.8)	<0.001
Number of drains placed	1 (1-1)	1 (1-1)	2 (1-2)	<0.001
Laboratory measures (blood)				
CRP (mg/L)	96 (39-173)	85 (32-165)	141 (95-190)	<0.001
Leukocytes (x 10^9/L)	15.7 (11.8-20.1)	14.8 (11.3-19.0)	20.1 (16.3-23.6)	<0.001
Hemoglobin (mmol/L)	6.6 (5.7-7.5)	6.8 (5.8-7.6)	6.0 (5.2-6.8)	<0.001
Thrombocytes (x 10^9/L)	351 (262-495)	329 (252-452)	540 (381-714)	<0.001
Laboratory measures (CSF)				
Leukocytes (x100/uL)	1.9 (0.3-5.7)	1.4 (0.2-4.3)	10.4 (2.5-53.1)	<0.001
Erythrocytes (x 10000/uL)	1.6 (0.2-7.4)	1.2 (0.2-6.9)	2.4 (0.8-10.6)	0.006
Binary leukocytes (%)	152 (28.3)	91 (20.0)	61 (74.4)	<0.001
Percentage neutrophils	51.7 (33.1-74.0)	47.8 (31.3-66.0)	85.0 (70.0-91.5)	<0.001
Neutrophil count (x 100/uL)	0.8 (0.1-4.9)	0.4 (0.0-2.3)	6.3 (0.6-38.0)	<0.001
Glucose (mmol/L)	3.4 (2.7-4.1)	3.5 (2.9-4.2)	2.3 (1.1-3.3)	<0.001
Total protein (g/L)	1.7 (0.8-2.8)	1.7 (0.8-2.7)	1.8 (1.1-3.3)	0.027
Culture results				
CSF and/or drain culture (%)	106 (19.7)	45 (9.9)	61 (74.4)	<0.001
Antibiotic use				
Any antibiotics started >4 days (%)	271 (50.5)	193 (42.4)	78 (95.1)	<0.001
Any empiric antibiotic therapy (%)	123 (22.9)	61 (13.4)	62 (75.6)	<0.001
Number of antibiotic started	1.0 (0-3)	1 (0-2)	4 (3-6)	<0.001

[a] p-value using χ^2, student's *t* or Mann-Whitney U test where appropriate.

Number of missing values prior to imputation: Other HAI – 37.2%; CRP – 11.2%; Leukocytes (blood) – 8.4%; Hemoglobin – 6.1%; Thrombocytes – 11.2%; CSF leukocytes – 29.2%; CSF erythrocytes 29.1%; CSF glucose 30.7%; CSF protein 29.2%; Culture (CSF and/or drain) – 19.9%. All others: no missing values.

Abbreviations: DRM – drain-related meningitis; IQR – interquartile range; ICU – intensive care unit; CSF – cerebrospinal fluid; EVD – external ventricular drain; HAI – Healthcare-associated infection; SAH – subarachnoid hemorrhage; IVH – intraventricular hemorrhage.

hemorrhage or (cerebellar) infarction and almost 60% (n = 312) of patients were admitted to the ICU during part of their stay. Patients were admitted for a median of 21 days (including readmissions within 30 days).

Table 2 | Outcome of backward stepwise logistic regression predicting the risk of drain-related meningitis.

Predictor	OR	95% CI	p-value
Drain type (EVD)	5.26	1.57-17.60	0.003
Number of drains placed	2.04	1.22-3.41	0.005
CRP	1.02	0.99-1.05	0.245
$(CRP/10)^2$	0.99	0.97-1.00	0.048
$(CRP/100)^3$	1.26	0.99-1.60	0.044
Leukocytes (blood)	1.08	1.01-1.16	0.018
Leukocytes (CSF)	1.42	1.15-1.75	< 0.001
Positive culture (drain and/or CSF)	13.70	5.58-33.62	< 0.001
Any empiric antibiotics started	4.33	1.79-10.5	< 0.001
Number of antibiotics started	1.32	1.04-1.68	0.021

Notes: Outcome of backward stepwise logistic regression, cut-off for exclusion $p < 0.05$. Odds ratio and confidence intervals are after bootstrapping, p-values and predictor selection are prior to bootstrapping and shrinkage. Predictors not retained in model: indication for drain placement, duration of admission, total drainage duration, number of days in intensive care unit, CSF glucose, CSF total protein.
Abbreviations: CI – confidence interval, CRP – C-reactive protein, CSF – cerebrospinal fluid, EVD – external ventricular drain, OR – Odds ratio.

Based on the results of the univariate analysis, the following variables were selected for multivariate analysis: indication for drain placement, duration of admission, number of drains placed, total drainage duration, duration of ICU admission, CRP, blood leukocytes, CSF leukocytes, CSF glucose, CSF protein, culture result (CSF and/or drain), total number of antibiotics started during admission, and whether empiric antibiotic therapy for drain-related meningitis was initiated.

$$P(DRM) = \frac{1}{1 + e^{-LP}}$$

LP = -8.014 + 0.713*Number of drains + 1.659* if EVD + 0.017*(CRP/10) + -0.015 *$(CRP/10)^2$ + 0.233*$(CRP/100)^3$ + 0.077*leukocytes(blood) + 0.350*ln(CSF leukocytes) + 2.617*positive culture + 0.281*number antibiotics started + 1.466* empiric therapy .

Figure 2 | Prediction rule for the development of drain-related meningitis.

Abbreviations: P(DRM) – probability of drain-related meningitis; LP – linear predictor; EVD – external ventricular drain; CRP – C-reactive protein; CSF – cerebrospinal fluid.

Table 2 shows the predictors retained in the model and their associated p-values. Despite its high p-value ($p = 0.230$), the linear CRP term was kept in the model in order to allow the significant high-power terms to be included. The prediction rule can be used to calculate the probability of meningitis for each patient (**figure 2**). Discriminatory power of the model as determined by the area under the ROC curve was 0.970 (95% CI: 0.954-0.986). Calibration of the final model was good (**figure 3**).

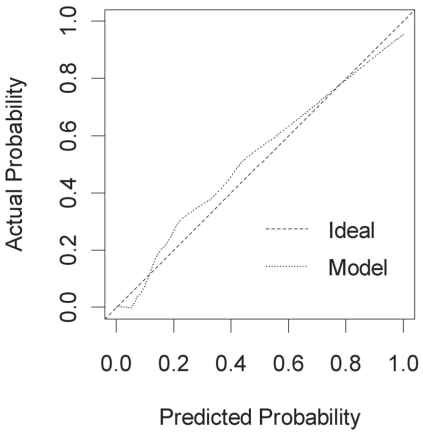


Figure 3 | Calibration plot of the model predicting drain-related meningitis.

The diagonal dashed line represents ideal prediction by the model, the pointed line predicted probabilities. Calibration, or the concordance between predicted and observed probability of infection, is adequate.

A cut-off in predicted probability of 0.107 resulted in 98.8% sensitivity, specificity of 87.9% and positive and negative predictive values of respectively 59.6% and 99.8% (**table 3**). The only missed infection was an infection with a coagulase-negative staphylococcus for which no antibiotics were started during admission and of which the patient recovered spontaneously. Selecting a cut-off probability of 0.175 missed three additional infections (sensitivity 95.1%), but only slightly improved specificity (91.0%) and positive predictive value (65.5%).

Table 3 | Two-by-two contingency table for predicted probability (*P(DRM)*) in relation to drain-related meningitis.

Predicted probability	DRM			Sensitivity	Specificity	PPV	NPV
	Yes	**No**	**Total**	**(%)**	**(%)**	**(%)**	**(%)**
P(DRM) > 0.107	81	55	136	98.8	87.9	59.6	99.8
P(DRM) ≤ 0.107	1	400	401	(93.4-99.9)	(84.6-90.8)	(50.8-67.9)	(98.6-99.9)
P(DRM) > 0.175	78	41	119	95.1	91.0	66.5	99.0
P(DRM) ≤ 0.175	4	414	418	(88.0-98.7)	(88.0-93.5)	(56.3-74.0)	(97.6-99.7)
Total	82	455	537				

Notes: Two-by-two contingency table for predicted probability cut-offs 0.107 and 0.175 in determining the presence of drain-related meningitis with associated sensitivity, specificity, positive and negative predictive values and 95% confidence intervals.

Abbreviations: NPV – negative predictive value, PPV – positive predictive value, P(DRM): predicted probability of drain-related meningitis.

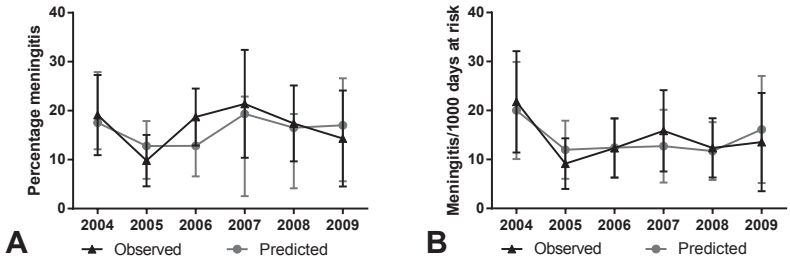


Figure 4 | Comparison of observed and predicted overall yearly infection rates.

Observed and predicted infection rates at the patient level (panel A) and expressed per 1000 drainage days at risk (panel B), including 95% confidence intervals. Predicted yearly infection rates are determined by the summed predicted probabilities and show good concordance with observed rates.

If a definite diagnosis is necessary at the patient level, application of the model reduced the number of charts to review manually to 25.3% (from 537 to 136 charts) while still identifying 98.8% of infections (81 out of 82). When interested in infection rates at the group level, the summed predicted probabilities reflect total infection percentages with good concordance (**figure 4**) and thus allow for surveillance without the need for manual confirmation.

DISCUSSION

The results of this study show that information stored in clinical data warehouses can successfully be used to predict rates of DRM in patients receiving an external ventricular or lumbar drain. The combination of drain characteristics, microbiology and clinical chemistry results and antibiotic use achieved 98.8% sensitivity and 87.5% specificity in detecting drain-related meningitis when applying a predicted probability cut-off of 0.107. Negative and positive predictive values were 99.8% and 56.9% respectively. Performing chart review only for those patients identified by the model to have DRM would reduce the number of manual chart reviews by 74.7%. Monitoring of longitudinal infection rates at the group level, however, can be performed without manual confirmation, thereby providing an efficient surveillance tool. This study can be viewed as proof-of-concept for use of regression model-based systems to perform surveillance of nosocomial infections at the group level.

As opposed to most detection models described previously, the model presented here uses data from a multitude of sources in a multivariable model, and case-finding is based on the weighted combination of predictors from each source. As opposed to classification

algorithms with case-finding based on broadly selected indicators, this weighted combination of predictors leads to high sensitivity for both culture-positive and culture-negative infections while maintaining acceptable positive predictive value. To the best of our knowledge, this is the first model also incorporating results of clinical chemistry and hematology analysis as predictors of healthcare-associated infections.

In comparison to conventional manual surveillance, this model reduces the time needed to perform surveillance, is less prone to error and less vulnerable to inter-rater variation. Furthermore, calculation of summed predicted probabilities for the at-risk population is an efficient surveillance tool to monitor changes in infection rates and determine when to perform in-depth analysis. Several studies have shown that although automated models using simplified and objective criteria may not always correctly predict absolute infection rates, such models may achieve reliable ranking of hospitals and accentuate differences between hospitals[12,27,28].

The large patient population included in this study allowed for the application of statistical methods as opposed to classification algorithms. Although the rule-of-thumb of ten events per predictor was violated, this does not necessarily lead to unreliable results[29]. Furthermore, the selected reference standard, the CDC/NHSN definition of healthcare-associated meningitis, has been measured consistently over time for purposes other than this research. Several other definitions of DRM have been used in literature[22,30-32], however they mostly require positive culture results and therefore have low sensitivity for culture-negative infections which occurred in twenty percent of cases in this population. Even though the CDC/NHSN definition is only partially applicable to neurosurgical patients who are comatose or sedated, the other definitions of DRM will also face this problem as many require the presence of clinical symptoms to confirm the diagnosis. Although it can be argued that healthcare-associated meningitis is a different clinical entity than meningitis secondary to cerebrospinal fluid drainage, the selected reference standard has been measured consistently and reliably over the six-year period and contains many similarities to other definitions proposed for drain-related meningitis. Imputation of missing values was used to prevent the introduction of bias in deriving the model. Since it is not possible to impute missing values for individual patients, a probability of infection cannot be computed for future patients with missing data. Out of the patients with an infection, only one had missing data for one predictor (CSF leukocyte count), thereby making underestimation of infection rates unlikely. Furthermore, predictors were only included if commonly determined in clinical practice. For this reason, parameters that have been described previously such as CSF lactate levels[33,34], CSF cytokine levels[23,35], and procalcitonin levels[36,37] were not considered for inclusion. The calculation of the cell-

index was considered as a tool to correct for blood-contaminated CSF[38]; however, since this measure could not be calculated in 65.3% of patients due to missing data, it was not included in the analyses. Finally, this model does not investigate infections occurring after discharge unless the patient is readmitted. However, contrary to surgical site infections, post-discharge surveillance is not as relevant since patients often remain in the hospital for a number of days after removal of the drain and it is customary for patients to return to their primary hospital when complications occur. These patients are then re-included in surveillance if readmission occurs within 30 days of discharge.

In summary, the model developed can accurately quantify rates of drain-related meningitis using multi-source data. The proposed model was developed using only retrospective data and although measures have been taken to prevent excessive optimism, prospective validation both within our centre and on a larger scale is necessary to assure applicability to other patient populations. This multivariable model-based approach can be applied to other types of nosocomial infections in the future. Also the development of methods to determine device utilization rates using data available through electronic healthcare records will further improve efficiency and reliability of surveillance.

Acknowledgements

The authors would like to thank all contributing data managers, H.E.M. Blok, H. den Breeijen and E. Mulder, for their contribution as well as the infection control professionals from the department of hospital hygiene and infection control (A.C.M. Gigengack-Baars, M.C.E. van der Jagt-Zwetsloot, S.M. van Dijk).

REFERENCE LIST

1. ECDC.Annual Epidemiological Report on Communicable Disease in Europe 2008. **2008**
2. Edwards JR, Peterson KD, Mu Y et al. National Healthcare Safety Network (NHSN) report: data summary for 2006 through 2008, issued December 2009. *Am J Infect Control.* **2009**;37:783-805.
3. Van der Kooi TI, Mannien J, Wille JC, van Benthem BH. Prevalence of nosocomial infections in The Netherlands, 2007-2008: results of the first four national studies. *J Hosp Infect.* **2010**;75:168-172.
4. Haley RW, Culver DH, White JW et al. The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals. *Am J Epidemiol.* **1985**;121:182-205.
5. Ehrenkranz NJ, Shultz JM, Richter EL. Recorded criteria as a "gold standard" for sensitivity and specificity estimates of surveillance of nosocomial infection: a novel method to measure job performance. *Infect Control Hosp Epidemiol.* **1995**;16:697-702.
6. Gastmeier P, Kampf G, Hauer T et al. Experience with two validation methods in a prevalence survey on nosocomial infections. *Infect Control Hosp Epidemiol.* **1998**;19:668-673.
7. Trick WE. Building a data warehouse for infection control. *Am J Infect Control.* **2008**;36:S75-81.
8. Bellini C, Petignat C, Francioli P et al. Comparison of automated strategies for surveillance of nosocomial bacteremia. *Infect Control Hosp Epidemiol.* **2007**;28:1030-1035.
9. Woeltje KF, Butler AM, Goris AJ et al. Automated surveillance for central line-associated bloodstream infection in intensive care units. *Infect Control Hosp Epidemiol.* **2008**;29:842-846.
10. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. *Clin Infect Dis.* **2009**;48:1268-1275.
11. Yokoe DS, Noskin GA, Cunnigham SM et al. Enhanced identification of postoperative infections among inpatients. *Emerg Infect Dis.* **2004**;10:1924-1930.
12. Platt R, Kleinman K, Thompson K et al. Using automated health plan data to assess infection risk from coronary artery bypass surgery. *Emerg Infect Dis.* **2002**;8:1433-1441.
13. Sands K, Vineyard G, Livingston J, Christiansen C, Platt R. Efficient identification of postdischarge surgical site infections: use of automated pharmacy dispensing information, administrative data, and medical record information. *J Infect Dis.* **1999**;179:434-441.
14. Bolon MK, Hooper D, Stevenson KB et al. Improved surveillance for surgical site infections after orthopedic implantation procedures: extending applications for automated data. *Clin Infect Dis.* **2009**;48:1223-1229.
15. Graham PL, III, San GP, Lutwick S, Haas J, Saiman L. Validation of a multicenter computer-based surveillance system for hospital-acquired bloodstream infections in neonatal intensive care departments. *Am J Infect Control.* **2004**;32:232-234.
16. Trick WE, Zagorski BM, Tokars JI et al. Computer algorithms to detect bloodstream infections. *Emerg Infect Dis.* **2004**;10:1612-1620.
17. Pokorny L, Rovira A, Martin-Baranera M, Gimeno C, onso-Tarres C, Vilarasau J. Automatic detection of patients with nosocomial infection by a computer-based surveillance system: a validation study in a general hospital. *Infect Control Hosp Epidemiol.* **2006**;27:500-503.
18. Brossette SE, Hacek DM, Gavin PJ et al. A laboratory-based, hospital-wide, electronic marker for nosocomial infection: the future of infection control surveillance? *Am J Clin Pathol.* **2006**;125:34-39.
19. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med.* **2007**;45:13-19.
20. Leverstein-Van Hall MA, Hopmans TEM, Van Der Sprenkel JWB et al. A bundle approach to reduce the incidence of external ventricular and lumbar drain-related infections: Clinical article. *J Neurosurg.* **2010**;112:345-353.
21. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control.* **2008**;36:309-332.
22. Lozier AP, Sciacca RR, Romagnoli MF, Connolly ES, Jr. Ventriculostomy-related infections: a critical review of the literature. *Neurosurgery.* **2002**;51:170-181.

23. Schade RP, Schinkel J, Roelandse FW et al. Lack of value of routine analysis of cerebrospinal fluid for prediction and diagnosis of external drainage-related bacterial meningitis. *J Neurosurg.* **2006**;104:101-108.
24. Pfisterer W, Muhlbauer M, Czech T, Reinprecht A. Early diagnosis of external ventricular drainage infection: results of a prospective study. *J Neurol Neurosurg Psychiatry.* **2003**;74:929-932.
25. Steyerberg EW. *Clinical Prediction Models: A practical approach to development, validation, and updating*. New York: Springer; **2008**.
26. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken: J. Wiley & Sons; **1987**.
27. Huang SS, Livingston JM, Rawson NS, Schmaltz S, Platt R. Developing algorithms for healthcare insurers to systematically monitor surgical site infection rates. *BMC Med Res Methodol.* **2007**;7:20.
28. Rubin MA, Mayer J, Greene T et al. An agent-based model for evaluating surveillance methods for catheter-related bloodstream infection. *AMIA Annu Symp Proc.* **2008**;631-635.
29. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol.* **2007**;165:710-718.
30. Mayhall CG, Archer NH, Lamb VA et al. Ventriculostomy-related infections. A prospective epidemiologic study. *N Engl J Med.* **1984**;310:553-559.
31. Sundbarg G, Nordstrom CH, Soderstrom S. Complications due to prolonged ventricular fluid pressure recording. *Br J Neurosurg.* **1988**;2:485-495.
32. Holloway KL, Barnes T, Choi S et al. Ventriculostomy infections: the effect of monitoring duration and catheter exchange in 584 patients. *J Neurosurg.* **1996**;85:419-424.
33. Leib SL, Boscacci R, Gratzl O, Zimmerli W. Predictive value of cerebrospinal fluid (CSF) lactate level versus CSF/blood glucose ratio for the diagnosis of bacterial meningitis following neurosurgery. *Clin Infect Dis.* **1999**;29:69-74.
34. Wong GK, Poon WS, Ip M. Use of ventricular cerebrospinal fluid lactate measurement to diagnose cerebrospinal fluid infection in patients with intraventricular haemorrhage. *J Clin Neurosci.* **2008**;15:654-655.
35. Lopez-Cortes LF, Marquez-Arbizu R, Jimenez-Jimenez LM et al. Cerebrospinal fluid tumor necrosis factor-alpha, interleukin-1beta, interleukin-6, and interleukin-8 as diagnostic markers of cerebrospinal fluid infection in neurosurgical patients. *Crit Care Med.* **2000**;28:215-219.
36. Berger C, Schwarz S, Schaebitz WR, Aschoff A, Schwab S. Serum procalcitonin in cerebral ventriculitis. *Crit Care Med.* **2002**;30:1778-1781.
37. Martinez R, Gaul C, Buchfelder M, Erbguth F, Tschaikowsky K. Serum procalcitonin monitoring for differential diagnosis of ventriculitis in adult intensive care patients. *Intensive Care Med.* **2002**;28:208-210.
38. Pfausler B, Beer R, Engelhardt K, Kemmler G, Mohsenipour I, Schmutzhard E. Cell index--a new parameter for the early diagnosis of ventriculostomy (external ventricular drainage)-related ventriculitis in patients with intraventricular hemorrhage? *Acta Neurochir (Wien).* **2004**;146:477-481.

3

Automated detection of healthcare associated infections: external validation and updating of a model for surveillance of drain-related meningitis

Maaike S.M. van Mourik[a], Karel G.M. Moons[b], Wouter W. van Solinge[c], Jan Willem Berkelbach van der Sprenkel[d], Luca Regli[d], Annet Troelstra[a], Marc J.M. Bonten[a, b]

[a] *Department of Medical Microbiology*
[b] *Julius Center for Health Sciences and Primary Care*
[c] *Department of Clinical Chemistry and Hematology*
[d] *Department of Neurosurgery, Rudolf Magnus Institute of Neuroscience*
All at University Medical Center Utrecht, Utrecht, The Netherlands

PLoS ONE 2012 7(12): e51509.

ABSTRACT

Objective – Automated surveillance of healthcare-associated infections can improve efficiency and reliability of surveillance. The aim was to validate and update a previously developed multivariable prediction model for the detection of drain-related meningitis (DRM).

Design – Retrospective cohort study using traditional surveillance by infection control professionals as reference standard.

Patients – Patients receiving an external cerebrospinal fluid drain, either ventricular (EVD) or lumbar (ELD) in a tertiary medical care center. Children, patients with simultaneous drains, < 1 day of follow-up or pre-existing meningitis were excluded leaving 105 patients in validation set (2010-2011) and 653 in updating set (2004-2011).

Methods – For validation, the original model was applied. Discrimination, classification and calibration were assessed. For updating, data from all available years was used to optimally re-estimate coefficients and determine whether extension with new predictors is necessary. The updated model was validated and adjusted for optimism (overfitting) using bootstrapping techniques.

Results – In model validation, the rate of DRM was 17.4/1000 days at risk. All cases were detected by the model. The area under the ROC curve was 0.951. The positive predictive value was 58.8% (95% CI 40.7-75.4) and calibration was good. The revised model also includes Gram stain results. Area under the ROC curve after correction for optimism was 0.963 (95% CI 0.953– 0.974). Group-level prediction was adequate.

Conclusions – The previously developed multivariable prediction model maintains discriminatory power and calibration in an independent patient population. The updated model incorporates all available data and performs well, also after elaborate adjustment for optimism.

INTRODUCTION

Surveillance and feedback of healthcare-associated infection (HAI) rates to healthcare workers is considered a cornerstone of infection prevention programs[1,2]. Policy makers and the public increasingly demand transparent reporting of infection rates to quantify quality of healthcare, for example through surveillance networks such as the National Healthcare Safety Network (NHSN) in the United States or the PREZIES network in the Netherlands[3-6]. Because of the potential impact of HAI rates on healthcare utilization and reimbursement, the development of efficient and reliable surveillance methods is of increasing importance. In many circumstances, manual chart review of all patients is still the only available method for surveillance, although it is prone to error due to effort dependent case-finding and the possibility of inconsistent interpretation of case definitions[7,8]. Possibilities for automated surveillance of HAI using a variety of data sources have been investigated over the past two decades with varying success[9].

A HAI for which routine surveillance is implemented in our institution is drain-related meningitis (DRM), a relatively frequent complication of the use of external ventricular (EVD) and lumbar (ELD) cerebrospinal fluid drains in neurosurgical patients. DRM rates range from 2 up to 25% per drain placed[10-12] or 7.5 to 32 infections per 1000 days at risk (DAR)[13-15]. Causative micro-organisms are often skin flora, such as coagulase-negative staphylococci and *Staphylococcus aureus*, although in some settings Gram-negative micro-organisms (eg enterobacteriaceae) play an important role[11,13]. Infection rates also depend on the definition applied. Since surveillance aims to generate insight into rates and characteristics of DRM, definitions are not necessarily identical to a clinical diagnosis entailing treatment consequences. Importantly, some case-definitions, including the CDC-definition for healthcare-associated meningitis, allow for diagnosis of an infection without the presence of bacterial growth from clinical cultures[16,17].

Recently, an accurate prediction model for the automated surveillance of DRM has been proposed which combines predictors from multiple sources to identify those patients which have a high probability of having developed DRM during their admission, both cases of DRM with and without documented pathogens in microbiological cultures (**figure 1**)[14]. Such a model can provide more timely and reliable rates of DRM and manual chart review can then be limited to high-risk patients (with a high predicted probability of DRM) while maintaining sensitivity of detection. Importantly, the predictors are all collected during routine clinical care which facilitates applicability of the model in practice[18].

$$P(DRM) = \frac{1}{1+e^{-LP}}$$

LP = -8.01 + 0.71*Number of drains + 1.66* if EVD + 0.017*(CRP/10) + -0.015 *$(CRP/10)^2$ + 0.23*$(CRP/100)^3$ + 0.08*leukocytes(blood) + 0.35*ln(CSF leukocytes) + 2.62*positive culture + 0.28*number antibiotics started + 1.47* empiric therapy .

Figure 1 | Previously derived prediction rule for drain-related meningitis.
For each individual patient, the model returns a predicted probability of DRM which can be used to classify patients.

Abbreviations: P(DRM) – probability of drain-related meningitis, LP – linear predictor, EVD – external ventricular drain, CRP – C-reactive protein, CSF – cerebrospinal fluid.

Prediction models require validation in independent patient populations to assess their validity and performance in future use[19,20]. This research presents the temporal validation of the DRM prediction model. Besides validation, optimal model performance in future patients can be achieved by updating the model using both derivation and validation data[21,22]. Several newly available predictors were also considered in model updating.

METHODS

Ethics statement

As described previously, use of anonymized data from the clinical data warehouse has been exempted from review by the institutional review board of our institution[23].

Development study details

For details on model development, please refer to[14]. In brief, logistic regression was used to develop a prediction model aimed at identifying patients that developed DRM after placement of an EVD or ELD. The study was conducted at University Medical Center Utrecht, a 1042-bed tertiary medical center. Patients who entered the routine surveillance performed by the department of hospital hygiene and infection control between January 1st 2004 and December 31st 2009 were included, with the exception of children, patients with less than one day of follow-up, patients with known meningitis at the time of placement of the first drain, patients admitted with a drain *in situ* or multiple simultaneous drains, military personnel and multiple (independent) admissions within the study period (n = 537 in analysis). All EVDs were placed in operating theatres and are tunneled five centimeters under the skin. Drains were not coated with antibiotics and all patients received perioperative antibiotic prophylaxis. ELDs were either inserted in the operating theatres or in sterile conditions on the neurology ward. Drains were not exchanged on a prophylactic basis and CSF samples were collected for culture and

biochemical analysis only when infection was clinically suspected. Clinical care data were obtained from the Utrecht Patient Oriented Database (UPOD), a clinical data warehouse developed for research purposes which links patient characteristics to results from clinical chemistry and medical microbiology laboratories and pharmacy records[23]. Missing data were imputed using multiple imputation, and internal validation was performed[24,25].

Outcome

As in model development, the outcome or reference standard was the development of DRM, which is defined as the occurrence of meningitis when the drain is *in situ* or within seven days of drain removal. Meningitis is defined according to the CDC-definition for healthcare-associated meningitis as applied by the department of hospital hygiene and infection control during routine manual surveillance. Presence of healthcare-associated meningitis requires either a positive culture or a combination of clinical signs, cerebrospinal fluid (CSF) analysis indicative of meningitis and initiation of empiric antimicrobial therapy by the physician. Importantly, this definition for meningitis allows for classification as a meningitis without bacterial growth from microbiological cultures and requires that cultures with skin flora are evaluated for possible contamination (**figure 2**)[14-16,26]. All charts were manually reviewed, and possible cases of infection were reviewed by at least two infection control professionals. In case of disagreement, consensus was reached through discussion.

NHSN/CDC definition of healthcare-associated meningitis for patients > 1 year of age
At least 1 of the following criteria:

(1) Patient has organism cultured from cerebrospinal fluid (CSF).

(2) Patient has at least one of the following symptoms with no other recognized cause: fever (>38°C), headache, stiff neck, meningeal signs, cranial nerve signs, or irritability.
And at least 1 of the following:
a. Increased white cells, elevated protein and/or decreased glucose in CSF.
b. Organisms seen on Gram stain of CSF.
c. Organisms cultured from blood.
d. Positive antigen test from blood, CSF or urine.
e. Diagnostic single antibody titer or 4-fold increase for pathogen.
And, if diagnosis is made antemortem, physician instituted appropriate antimicrobial therapy.

Modifications:
- Coagulase-negative staphylococci are classified as contamination if further clinical or biochemical abnormalities are absent.
- Blood-contaminated CSF is interpreted as positive if the leukocyte to erythrocyte ratio exceeds 1:100.

Figure 2 | Modified CDC-definition for healthcare-associated meningitis (reference standard).

Model validation and patient population

The previously developed model for the prediction of DRM was validated on an independent cohort of consecutive patients that received an EVD or ELD, selected from the same center though from a later time period (January 2010 to June 11th 2011), a so-called temporal validation[19,20]. In this time period, surveillance for ELDs was limited to drains placed in operating theatres in 2010 and discontinued in 2011. Data on device utilization is currently collected manually using electronic operating theatre and ICU records. Children (n = 13), patients with a meningitis at drain placement (11), patients who died within 24 hours (5) or who received multiple simultaneous drains (2) and those who were admitted with a drain already in situ (1) were excluded from the analysis, leaving 105 patients in the validation set. Approximately three-quarters of the EVD patients (75 of 99 patients) received a drain due to hydrocephalus after intracranial hemorrhage; nine percent received an EVD to treat increased intracranial pressure caused by a tumor. Five out of six ELD's were placed as a per-operative preventive measure. For each patient in the validation set, the reference standard was determined and predictor data was collected.

Predictors

Predictors were defined, collected and interpreted as in model development[14]. Predictors were selected for their ability to predict the development of DRM, irrespective of a causal association. Patient characteristics, administrative data (e.g., length of stay, ICU admissions) and the clinical parameters used in the prediction model (**figure 1**) were extracted from the clinical data warehouse. Besides the predictors obtained during model development, Gram stain results, the location to which the patient was discharged (i.e. deceased, home or other care facility) and urgency of admission as recorded in administrative files were also available. For each patient, all results obtained throughout the surveillance episode (duration of drainage plus seven days or up to discharge) were retrieved. For each predictor, the value most indicative of infection was used as parameter value; for example, for the peripheral blood leukocyte count the highest value measured during the surveillance episode was entered in the prediction rule.

Statistical analyses

Missing data were imputed using multiple imputation (10 iterations) to prevent bias that would have occurred if the analysis had been limited to complete cases only[24]. In **table S1**, a comparison of cases with and without missing data is given. For model validation, imputation was performed on the validation set only. A new imputation was run on the development and validation set combined for model update (see below). The original

model depicted in figure 1 was validated. Discrimination, classification and prediction at the group level (calibration-in-the-large) were assessed.

Model updating

Since datasets from both model development and validation were available, we investigated whether the original model could be improved or updated using both datasets combined and hence make maximal use of all data[21,27]. As opposed to model derivation, patients with multiple simultaneous drains were no longer excluded since they are not expected to be different in terms of diagnosis of DRM. Furthermore, during the updating process, one misclassification error in the model development data was resolved and all data were adapted accordingly thus slightly improving performance characteristics obtained in the model development. All predictors from the original model were included in the revised model to re-estimate their coefficients. In addition, the new predictors (Gram stain, urgency of admission and discharge destination) were added to the model if they significantly improved the model (likelihood ratio test, p-value of 0.05). Gram stain results were combined with CSF culture result and CRP was included as a fractional polynomial to accommodate the non-linear association between CRP and risk of DRM[28]. Estimates were derived from the 10 imputation sets and pooled using Rubin's rule, a method that takes into account variation within and between multiple imputation data sets[25].

Then, internal validation was performed by bootstrapping (100 samples per imputation set, including predictor selection using all predictors considered in model development and update) and a uniform shrinkage factor was applied, this to prevent over-optimism and to make the model generalizable to future patient populations[29]. The final model is presented along with its optimism-corrected performance characteristics. Analyses were performed with SPSS® version 19 (SPSS Inc, Chicago IL) and R version 2.14.1 (www.r-project.org).

RESULTS

Model validation

Model validation was performed on 105 patients who received 134 drains. Nearly all patients in the surveillance received an EVD (94.3%), due to discontinuation of ELD surveillance. The infection rate in the validation period was 17.3 per 1000 drainage days at risk (DAR). All infections occurred in patients receiving an EVD. In fifty percent of infections, no positive culture was obtained. Median age in the validation set was 59.3 years (model development 58.5 years), 65.7% of patients were female (model development 54.0%) and 71.5% received a drain to treat hydrocephalus after subarachnoid bleeding,

(intraventricular) hemorrhage or infarction (model development 49.0%) and in-hospital mortality after exclusion of patients who died within 24 hours of drain placement was 21.9%. The area under the ROC curve, which is a measure of discrimination, was 0.951 (95% confidence interval (CI) 0.914 to 0.988); during model development an area under the ROC curve of 0.976 (95% CI 0.965-0.987, without correction for optimism) was observed[14]. Calibration-in-the-large, a measure of the total number of infections in a specified time period, predicted 13.46 infections in 2010 (observed = 13 infections) and 6.06 infections between January 1st and June 10th 2011 (observed = 7). **Table 1** gives the contingency table obtained after application of the original prediction model and threshold.

Table 1 | Contingency table with results of model validation with 95% confidence intervals for sensitivity, specificity and predictive values.

Predicted probability	**DRM**				**(%)**	**95% CI**
	Yes	No	Total	Sensitivity	100.0	(83.2-100)
P(DRM) > 0.107	20	14	34	Specificity	83.5	(73.9-90.7)
P(DRM) ≤ 0.107	0	71	71	PPV	58.8	(40.7-75.4)
Total	20	85	105	NPV	100.0	(94.9-100)

Abbreviations: NPV – negative predictive value, PPV – positive predictive value, P(DRM) – predicted probability of drain-related meningitis.

Model update

The model was updated to incorporate newly available data and optimize performance in new patients. The total 2004-2011 dataset included 653 patients which received 863 drains. The observed infection rate was 14.1/1000 DAR (16.7/1000 DAR for EVDs, 6.0/1000 DAR for ELDs). Baseline characteristics and the results of model re-estimation are presented in **table 2**. Patients who developed DRM received multiple courses of antibiotics during their surveillance episode; most likely they suffered from or were suspected of other concomitant infections. The higher mortality in the non-affected group is in part caused by the shorter duration of follow-up in the deceased patients; hence they had less time to develop a DRM.

Table 2 | Model update results for 2004-2011 data, including baseline characteristics and results of univariable and multivariable analysis.

	Results of univariable analysis			Results of multivariable analysis[b]		
	no DRM	DRM	p-value[a]	Estimate	OR	95% CI
Median (IQR) or n (%)	**n = 549**	**n = 104**				
Baseline characteristics:						
Age (years)	59.3 (47.3-69.3)	56.1 (47.4-66.4)	0.599			
Sex (% female)	307 (55.9)	58 (55.8)	0.997			
In-hospital mortality (%)	105 (19.5)	12 (11.5)	0.064			
Duration of admission (days)	19 (11-30)	41 (29-63)	<0.001			
ICU admission (%)	322 (58.7)	75 (72.1)	0.010			
Predictors in previous model:						
Drain Type (% EVD)	352 (64.1)	93 (89.4)	<0.001	1.49	4.421	1.461-13.373
Number of drains placed	1 (1-1)	2 (1-2)	<0.001	0.52	1.687	1.154-2.698
CRP (mg/L)[c]	99 (37-183)	143 (94-189)	<0.001	-0.08	0.926	0.883-0.972
Peripheral leukocytes (x 10^9/L)	15.3 (11.4-19.4)	20.3 (16.4-24.3)	<0.001	0.08	1.090	1.022-1.153
CSF leukocytes (x100/uL)[d]	1.9 (0.2-6.4)	12.9 (2.7-83.8)	<0.001	0.20	1.224	1.058-1.416
CSF and/or drain culture[e] (%)	54 (9.8)	77 (74.0)	<0.001			
Any empiric antibiotic therapy (%)	72 (13.1)	81 (77.9)	<0.001	1.80	6.067	2.632-13.983
Number of antibiotics started	1 (0-2)	4 (3-6)	<0.001	0.20	1.225	0.988-1.519
New variables considered:						
Emergency admission (%)	312 (56.9)	68 (65.4)	0.109			
Discharge to			<0.001			
-Home	235 (42.8)	25 (24.0)				
-Other (deceased, care facility)	314 (57.2)	79 (76.0)				
CSF and/or drain culture or Gram stain[e]	59 (10.7)	79 (76.0)	<0.001	2.50	12.117	5.202-28.225

[a] p-value in univariable analysis by student's t test, Mann-Whitney U test or Chi-square where appropriate

[b] Results of the multivariable analysis are after bootstrapping (shrinkage factor 0.79). The intercept of the model was estimated at -6.615.

[c] In the multivariable analysis, all CRP values were divided by factor 10.

[d] In the multivariable analysis, CSF leukocytes were log transformed.

[e] Culture results corrected for contamination with skin flora; if no antibiotics were started, culture was classified as negative.

Abbreviations: CRP – C-reactive protein; CSF – cerebrospinal fluid; DRM – Drain-related meningitis; EVD – external ventricular drain; OR – Odd's ratio.

Table 3 | Model classification results with different predicted probability cut-offs. With increasing cut-off, the sensitivity decrease is associated with a decrease in number of charts requiring manual review for confirmation of infection.

P(DRM) cut-off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Charts to review (% of total)
0.025	100.0	62.5	33.5	100.0	310 (47.5)
0.050	100.0	78.3	46.6	100.0	223 (34.2)
0.075	99.0	82.7	52.0	99.8	198 (30.3)
0.100	99.0	85.8	56.9	99.8	181 (27.7)
0.125	99.0	86.7	58.5	99.8	176 (27.0)
0.150	98.1	87.6	60.0	99.6	170 (26.0)
0.175	95.2	89.1	62.3	99.0	159 (24.3)
0.200	93.3	90.0	63.8	98.6	152 (23.3)
0.225	90.4	91.1	65.7	98.0	143 (21.9)
0.250	87.5	91.8	66.9	97.5	136 (20.8)
0.275	83.7	93.4	70.7	96.8	123 (18.8)
0.300	82.7	94.2	72.9	96.6	118 (18.1)

Abbreviations: DRM – drain-related meningitis, NPV – negative predictive value, PPV – positive predictive value, P(DRM) – predicted probability for drain-related meningitis.

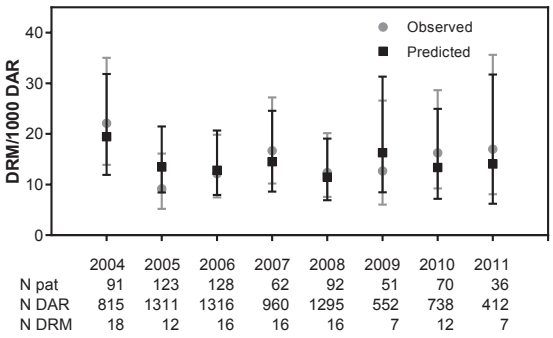


Figure 3 | Observed and predicted group-level infection rates using updated model, per 1000 days at risk with 95% confidence intervals.

Abbreviations: DRM – drain-related meningitis, DAR – days at risk, N pat – number of patients, N DAR – number of days at risk, N DRM – number of cases of drain-related meningitis.

In the multivariable analysis, using a fractional polynomial to fit the model to the CRP levels did not lead to the inclusion of higher power terms in the model and only the linear term was retained, albeit with a reversed direction. This is most likely because patients with a very high CRP level suffered from a different infection than DRM. The area under the ROC curve of the updated model was 0.972 before correction for optimism and 0.963 (95% CI 0.953-0.974) after correction for over-optimism. **Table 3** shows classification results with varying predicted probability cut-offs. Finally, yearly infection rates can be estimated by summing predicted probabilities (calibration-in-the-large) for all patients in each year group (**figure 3**).

DISCUSSION

The results of the present study demonstrate that the previously proposed model for the surveillance of DRM, in unaltered form, maintains its high discriminatory power and adequate group-level prediction in a new patient population from the same center. Patients included in the validation set were on average more seriously ill than in the derivation set, probably due to the discontinuation of surveillance of patients receiving ELDs. However, this did not impact model performance. Model update was performed to include predictors that recently became available and optimize the model. Performance of the updated model is similar to the original model. As described in **table 3**, a choice needs to be made between sensitivity and specificity in selecting a predicted probability cut-off; increasing the predicted probability cut-off reduces the number of charts to review at the cost of sensitivity. Since this model is applied to patients retrospectively and does not affect clinical decision making, it is worthwhile to, accept a sensitivity of 95.2% as opposed to 100.0% which will reduce the workload for manual review from 223 to 159 charts. As in model development, longitudinal surveillance at the group level can be performed using this model. Comparison of the original and revised model regression coefficients yields similar directions and slightly more conservative magnitudes due to the more stringent model shrinkage procedure used in the model update.

The observed rates of DRM in this study are in the upper part of the spectrum of rates published. The use of a broad definition that includes infections in which no micro-organisms were cultured from CSF may play a role (26% of the infections)[10]. Furthermore, as opposed to benchmarking data from Germany[30], both ICU and non-ICU patients are included in the surveillance and follow-up is extended beyond ICU discharge. The lower infection rate observed in patients who received an ELD (16.7 vs 6.0/1000 DAR) may be explained by the less severe underlying disease in these patients. Schade et al. also found lower DRM rates in patients receiving an ELD as compared to an EVD[31]; as in our population (data not shown), these patients often received an ELD for the prevention or treatment of CSF leakage. In other studies, a higher DRM rate was found in patients receiving ELDs which may be due to the inclusion of almost exclusively patients with underlying intracranial hemorrhage[15].

The updated model presented in this research is, to our knowledge, the only model developed to specifically survey the development of meningitis complicating the use of external CSF drains that has undergone temporal validation. Compared to other automated surveillance systems for (procedure-specific) HAI, this model is one of the few using data from multiple sources in a multivariable model which weights the individual predictors to generate a prediction. This is in contrast to the often seen binary classification

algorithms which use fewer data sources and often require positive cultures for case-finding[9,32,33]. In this model, positive cultures and antibiotic use are important predictors but no absolute requirement for the detection of infection, thus making it possible to identify those infections in which a positive culture was not obtained or for which the patient was not treated with standard empiric therapy. The study presented here confirms that this multivariable approach is valid for the surveillance of HAI, and may possibly be applied to other infections as well. Currently, use of the model requires extraction of predictor data from the electronic medical records and subsequent data processing prior to application of the prediction rule; ongoing developments in healthcare information technology are expected to facilitate the widespread implementation of such systems.

Since the number of external drains placed on a yearly basis is limited, the validation could only be performed on a relatively small patient population. Therefore, performing multiple imputation on this set of data required very relaxed settings which may cause unstable results. However, the model was subsequently revised and extended using the total population, one of the largest DRM cohorts to date, to make optimal use of available data and return the most reliable model possible. Although model update considered several new variables that have become available in the data warehouse, not all potential risk factors and diagnostic markers of DRM could be included. For example, there is no (field-defined) data on whether the drains were placed during an emergency procedure, how often drains were manipulated or whether there was cerebrospinal fluid leakage at the insertion site[10,11,34]. Markers of meningitis under investigation, such as procalcitonin and interleukins[35-37], are not routinely determined and thus not included. Furthermore, since the model is dependent on clinical practices, such as culture frequency and antibiotic use policies, the model may need to be adapted when implemented in new settings. However, the model will not be affected by differences in occurrence of causal risk factors assuming that clinical presentation and diagnostic workup remain unaffected. The effect on model performance of differences that may affect clinical presentation, such as use of antibiotic coated catheters, will need to be investigated further. When interpreting the results of this study, it must be realized that it has been developed for the purpose of infection surveillance after the fact, and not for realtime surveillance of infections. Several studies have attempted to identify parameters which can predict the onset of DRM, however with inconclusive results[37,38]. The current model could be used for more timely feedback of infection rates and may return more consistent results than manual surveillance.

This model for the surveillance of drain-related meningitis has now been temporally validated in a single center, and maintained performance despite small changes in case-mix of the validation set. Multi-center validation is currently ongoing to investigate

transportability to other hospitals and validity in patients with a different case-mix; also the effect of the use of antibiotic-coated catheters on model performance will be assessed. Several challenges still remain to achieve implementation in routine surveillance. Methods for handling of missing data in future patients need to be tested, and with the implementation in multiple centers, risk adjustment methods will be necessary to allow for valid comparison between centers. Another aspect that will require attention in the future is quantification of device utilization rates to generate infection rates with reliable numbers both in the numerator (this model) and the denominator.

Acknowledgements

The authors would like to thank H.E.M. Blok, H. den Breeijen and O. Cremer, for their contribution to data collection, as well as the infection control professionals from the department of hospital hygiene and infection control (M.C.E. van der Jagt-Zwetsloot, S.M. van Dijk).

SUPPLEMENTARY DATA

Table S1 | Comparison of patients with and without missing data. Complete cases have different underlying disease and are more likely to have developed DRM than non-complete cases.

Predictor	**complete case**[a]	**non-complete case**	
median or n(%)	**n = 410**	**n = 243**	**p-value**[b]
Age (years)	58.8	57.9	0.469
Sex (% female)	232 (56.6)	133 (54.7)	0.645
In-hospital death (%)	60 (14.6)	57 (23.5)	0.004
Duration of admission (days)	27	13	<0.001
Admission on ICU (%)	269 (65.6)	128 (52.7)	0.001
Indication for first drain (%)			<0.001
- SAH/IVH/infarction	259 (63.2)	101 (41.6)	
- CSF leakage	51 (12.4)	36 (14.8)	
- Per-operative	38 (9.3)	58 (23.9)	
- Trauma	12 (2.9)	4 (1.6)	
- Tumor	30 (7.3)	18 (7.4)	
- Other	38 (9.3)	26 (10.7)	
Emergency admission (%)	263 (64.1)	117 (48.3)	<0.001
DRM (%)	102 (24.9)	2 (0.8)	<0.001
Drain characteristics			
Drain type (% EVD)	304 (74.1)	141 (58.0)	<0.001
Total drain duration (days)	13	6	<0.001
Number of drains placed	1	1	<0.001
Laboratory measures			
CRP (mg/L)	132	85	<0.001
Leukocytes (x 10^9/L)	17.1	14.4	<0.001
CSF Leukocytes (x100/uL)	1.97	0.23	<0.001
CSF and/or drain culture positive (%)	119 (29.0)	8 (7.3)	<0.001
Gram stain positive for bacteria	71 (17.3)	3 (4.6)	0.009
Antibiotic use			
Any empiric antibiotic therapy (%)	148 (36.1)	5 (2.1)	<0.001
Number of antibiotic started	2	0	<0.001

[a] Data was termed complete if information for all model predictors was available.

[b] p-values by student's *t* test, Chi square or Mann-Whitney U test where appropriate

Abbreviations: ICU – intensive care unit, SAH – subarachnoid hemorrhage, IVH – intraventricular hemorrhage, CSF – cerebrospinal fluid, DRM – drain-related meningitis, EVD – external ventricular drain, CRP – C-reactive protein.

REFERENCE LIST

1. Haley RW, Culver DH, White JW et al. The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals. *Am J Epidemiol.* **1985**;121:182-205.
2. Gaynes R, Richards C, Edwards J et al. Feeding back surveillance data to prevent hospital-acquired infections. *Emerg Infect Dis.* **2001**;7:295-298.
3. Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med.* **2007**;357:1573-1575.
4. HAI Reporting Laws and Regulations. **2011**. APIC Last accessed: Feb 13 2012.
5. van der Kooi TI, Mannien J, Wille JC, van Benthem BH. Prevalence of nosocomial infections in The Netherlands, 2007-2008: results of the first four national studies. *J Hosp Infect.* **2010**;75:168-172.
6. Tokars JI, Richards C, Andrus M et al. The changing face of surveillance for health care-associated infections. *Clin Infect Dis.* **2004**;39:1347-1352.
7. Gastmeier P, Kampf G, Hauer T et al. Experience with two validation methods in a prevalence survey on nosocomial infections. *Infect Control Hosp Epidemiol.* **1998**;19:668-673.
8. Lin MY, Hota B, Khan YM et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. *JAMA.* **2010**;304:2035-2041.
9. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. *Clin Infect Dis.* **2009**;48:1268-1275.
10. Lozier AP, Sciacca RR, Romagnoli MF, Connolly ES, Jr. Ventriculostomy-related infections: a critical review of the literature. *Neurosurgery.* **2002**;51:170-181.
11. Chi H, Chang KY, Chang HC, Chiu NC, Huang FY. Infections associated with indwelling ventriculostomy catheters in a teaching hospital. *Int J Infect Dis.* **2010**;14:e216-e219.
12. Hoefnagel D, Dammers R, Ter Laak-Poort MP, Avezaat CJ. Risk factors for infections related to external ventricular drainage. *Acta Neurochir (Wien).* **2008**;150:209-214.
13. Arabi Y, Memish ZA, Balkhy HH et al. Ventriculostomy-associated infections: incidence and risk factors. *Am J Infect Control.* **2005**;33:137-143.
14. van Mourik MS, Groenwold RH, Berkelbach van der Sprenkel JW, van Solinge WW, Troelstra A, Bonten MJ. Automated detection of external ventricular and lumbar drain-related meningitis using laboratory and microbiology results and medication data. *PLoS One.* **2011**;6:e22846.
15. Scheithauer S, Burgel U, Ryang YM et al. Prospective surveillance of drain associated meningitis/ventriculitis in a neurosurgery and neurological intensive care unit. *J Neurol Neurosurg Psychiatry.* **2009**;80:1381-1385.
16. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control.* **2008**;36:309-332.
17. Holloway KL, Barnes T, Choi S et al. Ventriculostomy infections: the effect of monitoring duration and catheter exchange in 584 patients. *J Neurosurg.* **1996**;85:419-424.
18. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* **2009**;338:b375.
19. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* **2009**;338:b605.
20. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol.* **2008**;61:1085-1094.
21. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol.* **2008**;61:76-86.
22. Steyerberg EW. Updating for a new setting. In: Steyerberg EW, ed. *Clinical Prediction Models*. New York: Springer; 2009:361-90.
23. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med.* **2007**;45:13-19.

24. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* **2006**;59:1087-1091.
25. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken: J. Wiley & Sons; **1987**.
26. Leverstein-Van Hall MA, Hopmans TEM, Van Der Sprenkel JWB et al. A bundle approach to reduce the incidence of external ventricular and lumbar drain-related infections: Clinical article. *J Neurosurg.* **2010**;112:345-353.
27. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* **2004**;23:2567-2586.
28. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol.* **1999**;28:964-974.
29. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* **2001**;54:774-781.
30. KISS. ITS-KISS Reference data. **2012**
31. Schade RP, Schinkel J, Visser LG, Van Dijk JM, Voormolen JH, Kuijper EJ. Bacterial meningitis caused by the use of ventricular or lumbar cerebrospinal fluid catheters. *J Neurosurg.* **2005**;102:229-234.
32. Trick WE, Zagorski BM, Tokars JI et al. Computer algorithms to detect bloodstream infections. *Emerg Infect Dis.* **2004**;10:1612-1620.
33. Pokorny L, Rovira A, Martin-Baranera M, Gimeno C, onso-Tarres C, Vilarasau J. Automatic detection of patients with nosocomial infection by a computer-based surveillance system: a validation study in a general hospital. *Infect Control Hosp Epidemiol.* **2006**;27:500-503.
34. Korinek AM, Reina M, Boch AL, Rivera AO, De BD, Puybasset L. Prevention of external ventricular drain--related ventriculitis. *Acta Neurochir (Wien).* **2005**;147:39-45.
35. Martinez R, Gaul C, Buchfelder M, Erbguth F, Tschaikowsky K. Serum procalcitonin monitoring for differential diagnosis of ventriculitis in adult intensive care patients. *Intensive Care Med.* **2002**;28:208-210.
36. Lopez-Cortes LF, Cruz-Ruiz M, Gomez-Mateos J, Viciana-Fernandez P, Martinez-Marcos FJ, Pachon J. Interleukin-8 in cerebrospinal fluid from patients with meningitis of different etiologies: its possible role as neutrophil chemotactic factor. *J Infect Dis.* **1995**;172:581-584.
37. Schade RP, Schinkel J, Roelandse FW et al. Lack of value of routine analysis of cerebrospinal fluid for prediction and diagnosis of external drainage-related bacterial meningitis. *J Neurosurg.* **2006**;104:101-108.
38. Pfisterer W, Muhlbauer M, Czech T, Reinprecht A. Early diagnosis of external ventricular drainage infection: results of a prospective study. *J Neurol Neurosurg Psychiatry.* **2003**;74:929-932.

4

Automated surveillance for healthcare-associated infections: opportunities for improvement

Maaike S.M. van Mourik[a], Annet Troelstra[a], Wouter W. van Solinge[b], Karel G.M. Moons[c], Marc J.M. Bonten[a,c]

[a] *Department of Medical Microbiology*
[b] *Department of Clinical Chemistry and Hematology*
[c] *Julius Center for Health Sciences and Primary Care*
University Medical Center Utrecht, Utrecht, The Netherlands

Clinical Infectious Diseases 2013 Vol. 57(1): 85 - 93.
Published by Oxford University Press

ABSTRACT

Surveillance of healthcare-associated infections is a cornerstone of infection prevention programs and reporting of infection rates is increasingly required. Traditionally, surveillance is based on manual medical record review; however, this is very labor intensive and vulnerable to misclassification. Existing electronic surveillance systems based on classification algorithms using microbiology results, antibiotic use and/or discharge codes have increased the efficiency and completeness of surveillance by pre-selecting high-risk patients for manual review. However, shifting to electronic surveillance using multivariable prediction models based on available clinical patient data will allow for even more efficient detection of infection. With ongoing developments in healthcare information technology, implementation of the latter surveillance systems will become increasingly feasible. As with current predominantly manual methods, several challenges do remain such as completeness of post discharge surveillance and adequate adjustment for underlying patient characteristics, especially for comparison of healthcare-associated infections rates across institutions.

INTRODUCTION

Healthcare-associated infections (HAI) pose a considerable burden on medical care worldwide. Approximately 1.7 million people develop an HAI in United States hospitals each year and seven percent of patients admitted to European hospitals are affected by an HAI on any given day[1,2]. Over the past decades, HAI have more often been considered preventable complications of medical care and reporting of infection rates has become increasingly important not only for healthcare providers, but also to payers and the public. The ongoing extension of pay-for-performance programs and (mandatory) public reporting has increased the potential impact of HAI rates on healthcare reimbursement and public opinion[3-5]. Given the anticipated consequences of reported HAI rates, the accuracy of surveillance results is more of a concern than ever before.

Surveillance and feedback of HAI rates has been a cornerstone of infection prevention programs since the 1980's, as participation in surveillance programs was associated with reduced HAI incidence[6-8]. Surveillance programs require reliable HAI rates, delivered in a timely manner and as efficiently as possible. Traditionally, infection control professionals manually review medical records of all patients at risk for the specified HAI or conduct prospective inpatient surveillance. However, this is highly resource intensive and can therefore only be applied to limited patient populations. Moreover, case-definitions usually leave room for subjective interpretation and the completeness of case finding depends on the effort put in by surveyors[9,10]. Finally, differences between hospitals in how surveillance is implemented may affect reported infection rates. For instance, case finding limited to a single source of information (microbiology results, antibiotic prescriptions) may lead to underreporting[11,12], as does more thorough application of case-definitions[13].

These drawbacks of traditional HAI surveillance methods has led to initiatives to support or replace manual surveillance, either prospectively or retrospectively, by automated data collection from electronic medical records of microbiology results, antibiotic dispensing or administrative data[14-30]. Electronic surveillance methods aim to distinguish patients with high and low likelihoods of having developed an HAI in order to limit manual medical record review to high-risk patients; this improves efficiency while the number of missed HAI cases is maintained at an acceptable level. These electronic systems can be divided into two categories (**table 1**): classification algorithms that select high-risk patients based on the presence of indicators of infection and multivariable regression models that combine indicators of infection in a weighted formula to identify high-risk patients.

Table 1 | Overview of surveillance methodologies.

Method	Characteristics
Traditional surveillance	Manual medical record review of all patients at risk of specific HAI.
Classification algorithm	Limited number of electronic data sources in unweighted classification algorithm with consecutive dichotomous decision steps to identify high-risk patients for manual medical record review[a].
Multivariable regression model	Electronic data sources in a weighted, multivariable regression or prediction model to classify patients as low or high risk followed by manual review of high-risk patients[b].

Notes: [a] See figure 1A. [b] See figure 1B. Regression models may also be used to directly estimate rates of HAI at the group level. Abbreviations: HAI – healthcare-associated infection.

In this review, we will discuss that existing electronic surveillance systems based on classification algorithms are an improvement over manual surveillance, but that limitations remain with respect to the balance between completeness of case finding (sensitivity) and efficiency. More sophisticated multivariable prediction models, which to date have found only few applications in this context, address these limitations and may generate the high quality infection rates that are needed for (public) reporting and infection control research.

CLASSIFICATION ALGORITHMS

Electronic surveillance systems based on classification algorithms are analogous to a classification tree that classifies patients as low or high risk of having had an infection based on several consecutive dichotomous (yes/no) decision steps. In first instance, case finding is based on flagging patients with a certain dichotomous indicator of HAI, eg. positive microbiology culture results from the targeted infection site. In each subsequent step, classification of patients is further refined by applying a new 'HAI indicator' such as the 30-day postoperative interval, presence of a specified device or concurrent antimicrobial dispensing. Patients classified as high risk after these consecutive steps are then selected to undergo medical record review **(figure 1A)**[18,21,25-27]. This method has been applied most often to surgical site infections (SSI)[17-22] and (catheter-associated) bloodstream infections (CLABSI)[24-28], but also to ventilator-associated pneumonia, urinary tract infections or post-partum infections[28-30]. Most classification algorithms use one or two dichotomous indicators of infection stored in a structured format in hospital databases to flag high-risk patients, for instance microbiology culture results, (threshold) antimicrobial exposure or administrative data (discharge and procedure codes)[16-22,24,28-31]. Classification algorithms have also been used by payers (eg. insurance companies) to directly estimate HAI rates from claims data without manual confirmation and perform ranking of healthcare providers[31,32].

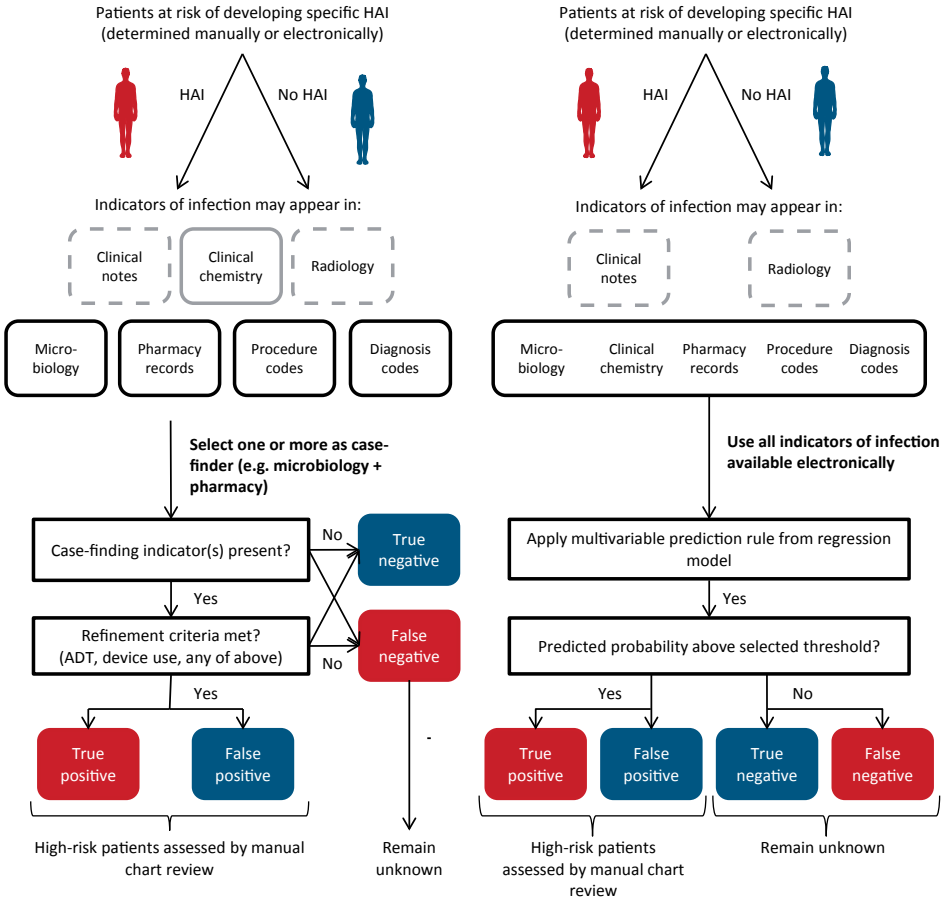


Figure 1 | Schematic overview of electronic surveillance system structures. In traditional surveillance all patients will undergo manual medical record review.

Panel A: Classification algorithm in which case finding is based on the (unweighted) presence of one or more indicators of infection followed by subsequent (dichotomous) refinement of flagged patients (based on administrative data, device presence, other indicators of infection). High-risk patients can then undergo medical record review. Panel B: Regression model in which all sources of data are included in the application of a weighted regression formula. All patients with a predicted probability of infection above the specified threshold will undergo manual medical record review to confirm the infection.
Red: Patient developed HAI, Blue: patient did not develop HAI. Sources of information circled in gray are not often used in the specified system. Sources of information with dashed lines are currently only available through natural language processing. Abbreviations: ADT –admission, discharge and transfer database, HAI – healthcare-associated infection.

As these classification algorithms aim to select, or rather exclude, patients that certainly did not have an HAI and do not require manual medical record review, the most important performance characteristics are sensitivity (high probability of detecting the truly affected patients) and negative predictive value, the probability that absence of a flag by the algorithm truly excludes HAI. The efficiency of the system can be assessed by the positive predictive value, the probability that the presence of a flag by the algorithm in fact represents HAI, or alternatively, by the number needed to screen or review, which is the number of medical records that must be manually assessed to identify one HAI[14]. In the setting of payer-based surveillance, important performance characteristics of classification algorithms are precision of the estimated total number of HAI at the group level and ranking accuracy[16,31].

Performance of existing classification algorithms

Compared to manual surveillance, electronic classification algorithms have a high efficiency, and some show improved inter-rater reliability and sensitivity, possibly due to increased consistency of classification of microbiology culture results and introduction of more systematic case finding strategies respectively[18,21,24,27]. **Table 2** lists performance characteristics of classifications algorithms for surveillance of HAI from recent studies and shows the gain in efficiency achieved.

However, the studies listed in **table 2** also illustrate some caveats of these classification algorithms. The fine balance between sensitivity and positive predictive value (efficiency) is exemplified by the SSI algorithms. In the algorithms with acceptable levels of sensitivity (> 80%), true infections were only confirmed in one of three to eight medical records flagged for review (number needed to screen 2.8-7.7); although, depending on the occurrence of the targeted HAI, this may still be a substantial workload reduction. Moreover, since not all types of HAI require positive microbiology results to meet case definitions[33], basing the initial case finding only on microbiology results will lead to lower sensitivity of detection. This is illustrated by the studies by Hollenbeak and Stamm and Betacchi which rely solely on microbiology results to identify SSI and intensive care-related infections respectively. The consequences of this restrictive case finding will depend on the type of HAI surveyed (BSI being the main exception) and clinical practice. Also, the use of discharge or procedure codes as a means to replace or support manual surveillance has revealed suboptimal results[34]. Although this readily available source of data may be an attractive alternative to costly manual surveillance, coding relies on an intermediate step of data interpretation where other priorities than those necessary for HAI surveillance may prevail.

Table 2 | Characteristics and performance of selected classification algorithms for the detection of healthcare-associated infections.

Study	Targeted HAI	n (% HAI)	Setting	Case finding based on	Refinement steps	Sens (%)	PPV (%)	NNS	Records to review (%)
Trick et al, 2004	CLABSI	135 (35.6)	Inpatient	Microbiology (blood culture)	Timing, other cultures, antibiotic use	81	62	1.3	47
Woeltje et al, 2008	CLABSI	771 (8.7)	ICU	Microbiology (blood culture)	Timing, CVC presence, other culture, fever	94	20	5.0	43
Woeltje et al, 2011	CLABSI	391 (22)	Inpatient (non-ICU)	Microbiology (blood culture)	Timing, CVC presence, other culture, fever	95	90	1.1	23
Pokorny et al, 2006[a]	BSI, VAP, CAUTI	194 (18)	ICU	2 or more of: microbiology, antibiotics or discharge code	Timing	94	56	1.8	30
Stamm et al 2012	CLABSI, VAP, CAUTI	141 HAI	ICU	Microbiology (NIM*)	Timing, ADT	67	39	2.6	unclear
Leth et al, 2010[b]	UTI	1513 (3.2)	Inpatient + post-discharge	Microbiology and/or antibiotic	Timing	77	93	1.1	2.6
	SSI	1513 (7.1)	Inpatient + post-discharge	Microbiology and/or antibiotic and/or discharge code and/or reoperation	Timing	72	71	1.4	14
Bolon et al., 2009[c]	SSI	6322 (1.7)	Inpatient	Antimicrobial exposure, diagnosis codes (index + readmission)	Timing	~90	25-40	2.5-4	4.0-6.2
Yokoe et al., 2012[c]	SSI	11159 (5.8)	Inpatient	Antimicrobial exposure, diagnosis codes (index + readmission)	Timing	89	18	5.6	18
Song et al., 2008	SSI	1226 (5.9)	Inpatient + outpatient	Antibiotic or readmission or discharge diagnosis (index, follow-up)	Timing	82	13	7.7	36
Hollenbeak et al., 2012[d]	SSI	1066 (8.8)	Inpatient	Microbiology (NIM*)	Timing, ADT	20	68	1.5	2.6
Gerbier et al., 2012	SSI	446 (8.5)	Inpatient	Antimicrobial exposure	Timing	68	34	2.9	17
				Microbiology	Timing	63	55	1.8	10
				Discharge code	Timing	26	83	1.2	38
				Combination of above	Timing	87	36	2.8	21

[a]Reference standard based on 194 patients; [b]Case finding for the reference standard includes patient questionnaire; [c]Data based on extrapolation of medical record review in a random sample (both studies use very similar methodology); [d]Reference standard from surgical quality improvement program (SQIP).
Notes: Reference standard was manual medical record review in routine surveillance unless specified otherwise. Sources: [18,20-23,25-28,30]. Abbreviations: ADT – admission, discharge, transfer system, BSI – Bloodstream infection, CAUTI – catheter-associated urinary tract infection, CLABSI – central line-associated bloodstream infection, HAI – healthcare-associated infection, ICU – Intensive care unit, NIM – Nosocomial infection marker, NNS – number needed to screen, PPV– positive predictive value, Sens – sensitivity, SSI – Surgical site infection, UTI – Urinary tract infection, VAP- Ventilator-associated pneumonia.

In short, electronic classification algorithms have important advantages over manual HAI surveillance both in terms of efficiency and systematic application of case-definitions, but suffer from vulnerability to restrictive case finding and the pay-off between sensitivity and efficiency.

MULTIVARIABLE REGRESSION MODELS

An alternative approach is the use of multivariable regression (prediction) models that combine indicators of infection (or predictors) simultaneously-rather than consecutively as in classification algorithms - to discriminate between high- and low-risk patients and select patients for manual medical record review **(figure 1B)**[17,32,35]. Given the patient's observed data, the weighted regression formula estimates the probability of having had an infection; this can then be used to stratify patients as high or low risk of HAI based on a probability threshold selected by the user[16,36].

This approach has been used to a limited extent by combining diagnosis and procedure codes extracted from claims databases with antimicrobial dispensing records[19,32]. As can be expected, models combining the largest number of predictors performed better than more parsimonious models[17]. Furthermore, the use of clinical patient data for the surveillance of a broad group of HAI in a simple model with five variables (antibiotic days, urine cultures performed, length of stay, wound cultures taken and age) achieved reasonable performance and could reduce the surveillance workload to 33 percent while maintaining sensitivity above eighty percent[35].

In **Table 3,** we have used empirical data to demonstrate the advantage of multivariable regression models over the use of classification algorithms for the surveillance of drain-related meningitis (DRM), a complication of external cerebrospinal fluid drainage in neurosurgical patients that is complex and cumbersome to diagnose manually (details in[36]). The case-definition does not necessitate isolation of micro-organisms from cerebrospinal fluid, accounting for the limited sensitivity of case finding based only on microbiology results (#1 & 2); as the population at risk has complex underlying disease, selection based on treatment with antibiotics leads to many false positive predictions (#3). Broadening the algorithm to combine both microbiology results and antimicrobial exposure clearly improved sensitivity, but at the cost of lower efficiency (#5). The multivariable regression model combining all available predictors of DRM simultaneously (#6) achieved near perfect sensitivity and had the highest efficiency, with high positive predictive value and low number needed to screen. Moreover, by changing probability thresholds based on

Table 3 | Empirical example of electronic surveillance based on multivariable regression compared to classification algorithms for drain-related meningitis.

	Model structure	Number of data sources	Sensitivity (%)	PPV (%)	NNS	Records to review (%)[a]
1	Microbiological culture	2 (Device use, microbiology)	80/104 (77)	80/164 (49)	2.0	164/653 (25)
2	Microbiological culture corrected for contaminant	3 (Device use, microbiology, pharmacy)	78/104 (75)	75/133 (56)	1.8	133/653 (20)
3	Any antimicrobial exposure (> 4 days)	2 (Device use, pharmacy)	100/104 (96)	100/331 (30)	3.3	331/653 (51)
4	Antimicrobial exposure to standard empiric regimen	2 (Device use, pharmacy)	81/104 (78)	81/153 (53)	1.9	153/653 (23)
5	2 OR 4	3 (Device use, microbiology, pharmacy)	102/104 (98)	102/235 (43)	2.3	235/653 (36)
6	Regression model[b]	4 (Device use, pharmacy, microbiology, clin chemistry)	102/104 (98)	102/170 (60)	1.7	170/653 (26)
7	Regression model (higher threshold)[c]	4 (Device use, pharmacy, microbiology, clin chemistry)	86/104 (83)	86/118 (73)	1.4	118/653 (18)

[a] Percentage medical records to review is defined as the number of medical records requiring manual review to confirm infection as a fraction of the entire population at risk that would require manual review if using traditional methods. [b] Model threshold = 0.15, [c] Model threshold = 0.3.
Notes: Data taken from [36]. All models presented require manual assessment of drain usage. If data was missing (models 1 to 4), results were classified as negative.
Abbreviations: NNS – number needed to screen, PPV – positive predictive value.

the number of missed cases deemed acceptable, users can modify the balance between sensitivity and specificity to their wishes (#7).

Although the input variables used by classification algorithms and regression models are similar, statistically weighted multivariable regression models have several advantages over classification algorithms.

- Multiple indicators are simultaneously taken into account, rather than consecutively, making case finding less restrictive and allowing for greater variation in clinical presentation of infection and thus higher detection rates (sensitivity).
- The weighted formula ensures higher positive predictive values than would be achieved if case finding were broadened in a non-weighted fashion (by combining indicators in an 'and/or' manner in classification algorithms).
- Prediction models offer flexibility with respect to the desired balance between sensitivity and efficiency. Increasing the predicted probability threshold above which patients are selected for manual review will lower the sensitivity of the model but increase efficiency and vice versa[36].

- Compared to previously developed regression models based on discharge and procedure codes extracted from claims data, a system which directly uses routinely collected clinical care data does not depend on data interpretation, such as assignment of discharge codes.
- Applications of prediction models may extend beyond their use as a stratification system. In particular, the number of infections at the group level can be estimated directly by summing predicted probabilities, thereby allowing estimation of infection rates without medical record review[36].

Developing multivariable prediction models

Increasing adoption of electronic medical records make multivariable prediction models feasible for the surveillance of HAI. Models can be developed using standard epidemiological methods and the number of data elements required may only be slightly larger than alternative classification algorithms with comparable sensitivity (**Table 3**)[37,38]. So-called clinical data warehouses are particularly suitable for this purpose; the drain-related meningitis model was developed on the Utrecht-Patient Oriented Database, a clinical data warehouse that links administrative and clinical databases[39]. Similar efforts and their application for HAI surveillance have been described for bloodstream infections[40] and other data mining efforts are currently being put towards the closely-related surveillance of other medical complications (eg. adverse drug events)[41,42]. With improvements in natural language processing, databases may in the future be enriched with increasingly detailed clinical information[42].

As with any attempt to develop surveillance methods, multivariable (logistic) regression models require data of a (sufficiently large) population at risk for the targeted HAI in which the candidate predictors have been measured and to which manual surveillance has been systematically and thoroughly applied by well-trained assessors (as reference or 'gold' standard). Definition and choice of candidate predictors should be driven by clinical knowledge and feasibility of data collection. Since the available data elements are often limited to those collected during routine clinical care, non-random missing data will be common, and selection of predictors which are only rarely determined may complicate implementation of the model. As missing values may also occur for indicators of infection that are commonly determined, methods such as multiple imputation of missing data may be needed to ensure reliable results during model development[43,44]. Performance of the resulting model can be assessed by the model's discrimination between infected and non-infected patients (eg. using the area under the ROC curve or, after introducing a threshold, the sensitivity, specificity and predictive values) and model calibration, which quantifies the concordance between predicted and observed probabilities of infection[37].

Applying so-called internal validation strategies will assist in generating realistic estimates of model performance when applied in future populations[38].

After model development, external validation in patient populations independent from the development sample in time and/or place is essential to assess the generalizability of the model and provide insight in how model performance is affected by different patient characteristics, diagnostic practices, and prescription patterns. Finally, ongoing monitoring of model quality and model updating is needed as changes in patient characteristics and clinical practice may deteriorate model accuracy over time[38].

REMAINING CHALLENGES

Even if adoption of automated HAI surveillance based on regression models will be attainable in the (near) future, several issues need to be addressed in order to achieve the high quality HAI rates needed by clinicians, policy makers and researchers; some are specific to electronic surveillance, both based on classification algorithms and regression models, methods, but many apply to all methods of surveillance. All methods of electronic surveillance require high quality data as well as more advanced programming skills, statistical methods and ongoing monitoring of data quality. A further challenge for all forms of electronic surveillance, is assessment of device presence, such as Foley catheters and central lines, as this determines the population at risk for device-related HAI. Depending on the functionality of the electronic medical record, the presence of devices is not always systematically recorded in a format that can easily be queried, necessitating manual assessment or, proxy-measures based on administrative data or, more promising, multivariable models[45-47]. Taking into account such functionality requirements during development and implementation of healthcare information systems would prevent this problem altogether.

Although methods for handling missing data in regression model development have been described extensively[43,48], only limited work has been done how to apply developed regression models to new patients in which not all predictors are available[49]. As the clinical situation will determine the likelihood of diagnostic testing, there will be missing values for some patients, in particular for laboratory and microbiological analysis. The best method for incorporating these (non-random) missing values in the context of HAI surveillance remains to be determined. Importantly, manual surveillance also relies on the performance of diagnostic testing and missing values also affect classification algorithms, as missing values for HAI indicators are usually classified as negative.

A challenge for all forms of surveillance, both manual and electronic, is achieving adequate postdischarge surveillance. Some HAI may occur after patient discharge, and even if patients do return to the initial hospital, such infections are prone to be missed unless attention to readmissions is explicitly incorporated into surveillance systems. The relevance of incomplete postdischarge surveillance on overall infection rates will depend on the type of HAI targeted (eg. length of follow-up) and on the likelihood that patients will return to the same healthcare provider. Although surveillance using claims data can more easily be extended over multiple care providers, discharge codes and other indirect measures are in turn more likely to suffer from misclassification[19]. In time however, automated methods may achieve higher sensitivity of detecting these infections through linkage across (re)admissions and facilities.

Most electronic systems developed so far, both classification algorithms and regression models, have focused on retrospective surveillance; that is, detection of infection when it has already occurred and treatment has been initiated. Specifically, initiation of antimicrobial therapy and bacterial growth several days after collection of patient samples often trigger the detection system. An appealing alternative would be real-time surveillance to detect onset of infection and ensure timely recognition. Development of such models, however, is even more challenging not only methodologically but also by limitation in the predictors available; by definition culture results are not yet available and the 'time of prediction' is aimed to precede start of antimicrobial treatment.

Due to the limitations of manual surveillance, both in terms of efficiency and reliability, research comparing alternative methods of surveillance is complicated by difficulties in obtaining a valid reference standard. Furthermore, partial and differential verification of the true infection status may bias results and, therefore, reported performance characteristics must be interpreted with care[50]. Moreover, all methods of surveillance described here rely - at least partly - on manual surveillance to confirm the presence of an infection. Minimizing subjective interpretation in case ascertainment is, therefore, needed to further improve HAI surveillance. For instance, more objective and quantifiable parameters have been defined for the detection of ventilator-associated conditions; this has been proposed as an alternative for surveillance of ventilator-associated pneumonia, an infection that is notoriously difficult to diagnose[51]. With ongoing improvements in electronic algorithms and a shift towards measuring more objective outcomes, electronic surveillance may become the mainstay for some HAI, at least for the purpose of quality assessment.

Finally, if the HAI rates are to be used for benchmarking or comparisons across multiple institutions, accurate adjustment for underlying risk (case-mix) is of vital importance. Development of more refined methods of case-mix adjustment is ongoing, and these efforts may benefit from the same improvements in healthcare information technology as the surveillance methods themselves[52].

CONCLUSION

The use of electronic classification algorithms for surveillance of healthcare-associated infections has increased reliability compared to the traditional surveillance through manual medical record review. However, use of more sophisticated multivariable prediction or regression models is likely to further improve in sensitivity and efficiency of surveillance programs. Some important challenges such as postdischarge surveillance, quantification of device utilization and case-mix adjustment are partly common to all methods of surveillance and need to be addressed in the future. With ongoing improvements in healthcare information technology, implementation of regression models will become feasible for widespread use and improve quality and capacity of surveillance.

Acknowledgements

We would like to thank the department of hospital hygiene and infection control for providing data on drain-related meningitis resulting from their ongoing surveillance efforts.

Funding

This work was supported by an Alexandre Suermann stipendium provided by the UMC Utrecht to MvM, the Netherlands Organization for Scientific Research (NWO VICI Grant 918.76.611) to MB and (NWO VICI and TOP Grant 9120.8004 and 918.10.615) to KGMM.

Conflict of interest

The authors have no conflicts of interest.

REFERENCE LIST

1. Klevens RM, Edwards JR, Richards CL, Jr. et al. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public Health Rep.* **2007**;122:160-166.
2. European Centre for Disease Prevention and Control. Annual Epidemiological Report on Communicable Disease in Europe 2008. **2008** Stockholm
3. Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med.* **2007**;357:1573-1575.
4. HAI Reporting Laws and Regulations. **2011**. APIC Last accessed: Feb 13 2012.
5. Health Protection Agency. Sixth report of the mandatory surveillance of surgical site infection in orthopaedic surgery, April 2004 to March 2010. **2010** London
6. Haley RW, Culver DH, White JW et al. The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals. *Am J Epidemiol.* **1985**;121:182-205.
7. Gaynes R, Richards C, Edwards J et al. Feeding back surveillance data to prevent hospital-acquired infections. *Emerg Infect Dis.* **2001**;7:295-298.
8. Geubbels EL, Nagelkerke NJ, Mintjes-De Groot AJ, Vandenbroucke-Grauls CM, Grobbee DE, De Boer AS. Reduced risk of surgical site infections through surveillance in a network. *Int J Qual Health Care.* **2006**;18:127-133.
9. Klompas M. Interobserver variability in ventilator-associated pneumonia surveillance. *Am J Infect Control.* **2010**;38:237-239.
10. Mayer J, Greene T, Howell J et al. Agreement in Classifying Bloodstream Infections Among Multiple Reviewers Conducting Surveillance. *Clin Infect Dis.* **2012**;55:364-370.
11. Gastmeier P, Brauer H, Hauer T, Schumacher M, Daschner F, Ruden H. How many nosocomial infections are missed if identification is restricted to patients with either microbiology reports or antibiotic administration? *Infect Control Hosp Epidemiol.* **1999**;20:124-127.
12. Glenister HM, Taylor LJ, Bartlett CL, Cooke EM, Sedgwick JA, Mackintosh CA. An evaluation of surveillance methods for detecting infections in hospital inpatients. *J Hosp Infect.* **1993**;23:229-242.
13. Klompas M. Eight initiatives that misleadingly lower ventilator-associated pneumonia rates. *Am J Infect Control.* **2012**;40:408-410.
14. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. *Clin Infect Dis.* **2009**;48:1268-1275.
15. Evans RS, Larsen RA, Burke JP et al. Computer surveillance of hospital-acquired infections and antibiotic use. *JAMA.* **1986**;256:1007-1011.
16. Platt R, Yokoe DS, Sands KE. Automated methods for surveillance of surgical site infections. *Emerg Infect Dis.* **2001**;7:212-216.
17. Sands K, Vineyard G, Livingston J, Christiansen C, Platt R. Efficient identification of postdischarge surgical site infections: use of automated pharmacy dispensing information, administrative data, and medical record information. *J Infect Dis.* **1999**;179:434-441.
18. Yokoe DS, Khan Y, Olsen MA et al. Enhanced surgical site infection surveillance following hysterectomy, vascular, and colorectal surgery. *Infect Control Hosp Epidemiol.* **2012**;33:768-773.
19. Miner AL, Sands KE, Yokoe DS et al. Enhanced identification of postoperative infections among outpatients. *Emerg Infect Dis.* **2004**;10:1931-1937.
20. Song X, Cosgrove SE, Pass MA, Perl TM. Using hospital claim data to monitor surgical site infections for inpatient procedures. *Am J Infect Control.* **2008**;36:S32-S36.
21. Bolon MK, Hooper D, Stevenson KB et al. Improved surveillance for surgical site infections after orthopedic implantation procedures: extending applications for automated data. *Clin Infect Dis.* **2009**;48:1223-1229.
22. Hollenbeak CS, Boltz MM, Nikkel LE, Schaefer E, Ortenzi G, Dillon PW. Electronic measures of surgical site infection: implications for estimating risks and costs. *Infect Control Hosp Epidemiol.* **2011**;32:784-790.

23. Gerbier-Colomban S, Bourjault M, Cetre JC, Baulieux J, Metzger MH. Evaluation study of different strategies for detecting surgical site infections using the hospital information system at lyon university hospital, france. *Ann Surg.* **2012**;255:896-900.
24. Lin MY, Hota B, Khan YM et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. *JAMA.* **2010**;304:2035-2041.
25. Woeltje KF, Butler AM, Goris AJ et al. Automated surveillance for central line-associated bloodstream infection in intensive care units. *Infect Control Hosp Epidemiol.* **2008**;29:842-846.
26. Woeltje KF, McMullen KM, Butler AM, Goris AJ, Doherty JA. Electronic surveillance for healthcare-associated central line-associated bloodstream infections outside the intensive care unit. *Infect Control Hosp Epidemiol.* **2011**;32:1086-1090.
27. Trick WE, Zagorski BM, Tokars JI et al. Computer algorithms to detect bloodstream infections. *Emerg Infect Dis.* **2004**;10:1612-1620.
28. Stamm AM, Bettacchi CJ. A comparison of 3 metrics to identify health care-associated infections. *Am J Infect Control.* **2012**;40:688-691.
29. Pokorny L, Rovira A, Martin-Baranera M, Gimeno C, onso-Tarres C, Vilarasau J. Automatic detection of patients with nosocomial infection by a computer-based surveillance system: a validation study in a general hospital. *Infect Control Hosp Epidemiol.* **2006**;27:500-503.
30. Leth RA, Norgaard M, Uldbjerg N, Thomsen RW, Moller JK. Surveillance of selected post-caesarean infections based on electronic registries: validation study including post-discharge infections. *J Hosp Infect.* **2010**;75:200-204.
31. Huang SS, Placzek H, Livingston J et al. Use of Medicare Claims to Rank Hospitals by Surgical Site Infection Risk following Coronary Artery Bypass Graft Surgery. *Infect Control Hosp Epidemiol.* **2011**;32:775-783.
32. Platt R, Kleinman K, Thompson K et al. Using automated health plan data to assess infection risk from coronary artery bypass surgery. *Emerg Infect Dis.* **2002**;8:1433-1441.
33. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control.* **2008**;36:309-332.
34. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. *Clin Infect Dis.* **2009**;49:949-955.
35. Broderick A, Mori M, Nettleman MD, Streed SA, Wenzel RP. Nosocomial infections: validation of surveillance and computer modeling to identify patients at risk. *Am J Epidemiol.* **1990**;131:734-742.
36. van Mourik MS, Moons KG, van Solinge WW et al. Automated Detection of Healthcare Associated Infections: External Validation and Updating of a Model for Surveillance of Drain-Related Meningitis. *PLoS ONE.* **2012**;7:e51509.
37. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* **2009**;338:b604.
38. Steyerberg EW. *Clinical Prediction Models: A practical approach to development, validation, and updating.* New York: Springer; **2009**.
39. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med.* **2007**;45:13-19.
40. Wisniewski MF, Kieszkowski P, Zagorski BM, Trick WE, Sommers M, Weinstein RA. Development of a clinical data warehouse for hospital infection control. *J Am Med Inform Assoc.* **2003**;10:454-462.
41. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. *J Am Med Inform Assoc.* **2003**;10:115-128.
42. Murff HJ, FitzHenry F, Matheny ME et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA.* **2011**;306:848-855.
43. Janssen KJ, Donders AR, Harrell FE, Jr. et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol.* **2010**;63:721-727.
44. Oostenbrink R, Moons KG, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: prospects and problems. *J Clin Epidemiol.* **2003**;56:501-506.
45. Wright MO, Fisher A, John M, Reynolds K, Peterson LR, Robicsek A. The electronic medical record as a tool for infection surveillance: successful automation of device-days. *Am J Infect Control.* **2009**;37:364-370.

46. Hota B, Harting B, Weinstein RA, Lyles RD, Bleasdale SC, Trick W. Electronic algorithmic prediction of central vascular catheter use. *Infect Control Hosp Epidemiol.* **2010**;31:4-11.

47. Meddings J, Saint S, McMahon LF, Jr. Hospital-acquired catheter-associated urinary tract infection: documentation and coding issues may reduce financial impact of Medicare's new payment policy. *Infect Control Hosp Epidemiol.* **2010**;31:627-633.

48. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* **2006**;59:1087-1091.

49. Janssen KJ, Vergouwe Y, Donders AR et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem.* **2009**;55:994-1001.

50. de Groot JA, Bossuyt PM, Reitsma JB et al. Verification problems in diagnostic accuracy studies: consequences and solutions. *BMJ.* **2011**;343:d4770.

51. Klompas M, Khan Y, Kleinman K et al. Multicenter Evaluation of a Novel Surveillance Paradigm for Complications of Mechanical Ventilation. *PLoS One.* **2011**;6:e18062.

52. Mu Y, Edwards JR, Horan TC, Berrios-Torres SI, Fridkin SK. Improving risk-adjusted measures of surgical site infection for the national healthcare safety network. *Infect Control Hosp Epidemiol.* **2011**;32:970-986.

5

Validation of an automated surveillance approach for drain-related meningitis: a multicenter study

Maaike S.M. van Mourik[a], Annet Troelstra[a], Jan Willem Berkelbach van der Sprenkel[b], Marischka C.E. van der Jagt-Zwetsloot[c], Jolande H. Nelson[d], Piet Vos[e], Mark P. Arts[f], Paul J.W. Dennesen[g], Karel G.M. Moons[h], Marc J.M. Bonten[a, h]

[a] *Department of Medical Microbiology, University Medical Center Utrecht*
[b] *Department of Neurosurgery, University Medical Center Utrecht*
[c] *Department of Hospital Hygiene and Infection Prevention, University Medical Center Utrecht*
[d] *Department of Infection Control and Prevention, St. Elisabeth Ziekenhuis, Tilburg*
[e] *Department of Intensive Care, St. Elisabeth Ziekenhuis, Tilburg*
[f] *Department of Neurosurgery, Medisch Centrum Haaglanden, The Hague*
[g] *Department of Intensive Care, Medisch Centrum Haaglanden, The Hague*
[h] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht*

Manuscript under review

ABSTRACT

Objective – Manual surveillance of healthcare-associated infections is cumbersome and vulnerable to subjective interpretation. Automated systems are under development to improve efficiency and reliability of surveillance, for example by selecting high-risk patients requiring manual chart review. The aim was to externally validate a previously developed multivariable prediction modeling approach for detecting drain-related meningitis (DRM) in neurosurgical patients and assess its merits compared to conventional methods of automated surveillance.

Methods – Prospective cohort study in three hospitals assessing the accuracy and efficiency of two automated surveillance methods - the multivariable prediction model and a classification algorithm - for detecting DRM, using surveillance by manual chart review as reference standard. All were performed independently of each other. Patients receiving cerebrospinal fluid drains were included (2012 – 2013), except children, patients deceased within 24 hours or with pre-existing meningitis. Data required by automated surveillance methods were extracted from routine care clinical data warehouses.

Results – In total, DRM occurred in 37 of 366 external cerebrospinal fluid drainage episodes (12.3/1000 drain days at risk). The multivariable prediction model had good discriminatory power (area under the ROC curve 0.91 – 1.00 by hospital), adequate overall calibration and could identify high-risk patients requiring manual confirmation with 97.3% sensitivity and 52.2% positive predictive value, decreasing the workload for manual surveillance by 81%. The multivariable approach was more efficient than classification algorithms in 2 of 3 hospitals.

Conclusions – When applied in multiple hospitals, automated surveillance of DRM using a multivariable prediction model considerably reduced the burden for manual chart review at near-perfect sensitivity.

INTRODUCTION

Electronically collected routine care data is increasingly employed to meet growing demands for reliable and timely information on healthcare-associated infection (HAI) rates[1-3]. For several decades, surveillance and feedback of HAI rates, for example within national networks, has been a fundamental component of infection prevention programs[4-7]. Traditionally, surveillance is performed by infection preventionists who manually review patient charts for the occurrence of targeted HAIs. This approach, however, is known to be labor-intensive, effort-dependent and vulnerable to subjective interpretation[8-11]. Expansion of surveillance volume requirements and public reporting of HAI rates has stimulated the use of (semi-) automated systems that combine various sources of data captured in electronic health records (EHRs) to support or replace manual surveillance[12,13].

Most of these automated surveillance systems aim to classify patients by their likelihood of having developed the targeted HAI and thereby restrict manual chart review to high-risk patients[2,12]. Data stored in EHRs can be harnessed in several ways to achieve this stratification. Using drain-related meningitis (DRM) as a prototype infection, a system based on a multivariable regression model was recently developed that demonstrated a favorable balance between sensitivity of DRM detection and workload reduction[2,14].

Drain-related meningitis, or ventriculitis or ventriculostomy-related infection, is an infection of the meninges and/or ventricles complicating the use of external cerebrospinal fluid (CSF) drains, in particular ventricular (EVD) and lumbar (ELD) drains[15-18]. These drains are placed to manage increased intracranial pressure, for example due to intracerebral hemorrhage or trauma, as per-operative prophylaxis or as treatment for cerebrospinal fluid leakage. DRM has been described to occur in up to 22% of drainage episodes[15-17]. Surveillance of DRM is complicated by a case-definition that is not straightforward and does not require a positive microbiological culture for diagnosis[18,19]; in addition, heterogeneous underlying disease and sometimes extensive comorbidities in this patient population hinder unequivocal application of diagnostic criteria.

The previously developed automated surveillance system based on multivariable regression modeling could accurately detect patients who developed DRM as a consequence of external CSF drainage. As prediction models require validation in both time and place in order to establish their transportability to other settings, this study presents a prospective, external validation of the multivariable automated surveillance approach to DRM in comparison to traditional manual methods[20,21]. In addition, it assesses the merits of the multivariable approach in relation to an alternative automated surveillance strategy based

on a classification algorithm and evaluates whether differences in performance can be linked to hospital or patient characteristics.

METHODS

Study design

This multicenter prospective cohort study compares two methods of automated surveillance to traditional, manual chart review (reference standard). All methods of surveillance were implemented independently of each other. Importantly, all surveillance approaches aim to identify those patients that have developed DRM at some time during their course of admission, and explicitly do not prospectively predict the onset of DRM. The institutional review boards of participating hospitals reviewed the study protocol and waived the requirement for informed consent.

Study population: hospitals and participants

All Dutch hospitals performing neurosurgical procedures were invited to participate in this study (n = 16). In addition to the University Medical Center Utrecht, a 1042-bed academic medical center where initial development of the model took place, two large non-academic teaching hospitals volunteered: St. Elisabeth Hospital in Tilburg (543 beds) and Medisch Centrum Haaglanden in The Hague (673 beds). Data were collected between January 1st 2012 and December 31st 2012 in the first two hospitals and between April 15th 2012 and April 30th 2013 in the latter. Hospitals were randomly assigned the letters A, B and C for further reference.

All patients receiving one or more EVDs or ELDs were eligible for inclusion and were identified by either routine infection control practices (hospitals B, C) or by a specifically developed electronic inclusion form filled in by neurosurgery and intensive care staff (hospital A). As in model development, children (below the age of 18), patients with pre-existing meningitis or who died or were discharged within 24 hours of drain placement and patients transferred from elsewhere with a drain *in situ* were excluded.

At the beginning of the study, infection control staff and treating physicians from each hospital completed a questionnaire on clinical practices regarding drain placement and maintenance, diagnosis and treatment of (suspected) drain-related meningitis and methods of DRM surveillance (**table 1**).

Table 1 | Characteristics of drain placement, diagnosis, treatment and surveillance of drain-related meningitis - by hospital.

	Hospital A	**Hospital B**	**Hospital C**
Drain placement			
Location of EVD placement	Operating theaters	Operating theaters	Operating theaters
- Placed by	- Neurosurgeon, residents	- Neurosurgeon, residents	- Neurosurgeon, residents
Location of ELD placement	Treatment room, bedside	Operating theaters, treatment room, bedside	Operating theaters, treatment room
- Placed by	- Residents, nurse-practitioners	- Neurosurgeon, residents, other	- Neurosurgeon, residents, other
Use of antibiotic-coated EVD	Sometimes (always > month 6)	Yes (always)	Starting month 10 of study
Per-operative antibiotic prophylaxis (EVD)	Yes	Yes	Yes
Per-operative antibiotic prophylaxis (ELD)	No	Sometimes	Sometimes
Drain maintenance			
Routine exchange drain	No	No	No
Wards providing EVD care	ICU, Step-down, ward	ICU, Step-down, ward	ICU, Step-down
Wards providing ELD care	ICU, Step-down, ward	ICU, Step-down, ward	ICU, Step-down, ward
Diagnosis			
Frequency of CSF sampling	Daily	Indication-only	Indication-only
Location of CSF sampling	Proximal	Proximal or CSF collection bag	Proximal
Routine culturing of drain tip	No (indication-only)	No (indication-only)	No (indication-only)
Treatment			
Empiric treatment regimen	Vancomycin + (ceftazidim or meropenem)	Flucloxacillin + ceftazidim	Vancomycin + ceftazidim or Flucloxacillin + ceftriaxone
Surveillance (ref standard)			
Performed by	Research team	Infection control professional	Infection control professional
Second reviewer confirmation?	Yes, physician	Yes, neurosurgeon or physician	Yes, infection preventionist

Abbreviations: CSF – cerebrospinal fluid, ELD – external lumbar drain, EVD – external ventricular drain, ICU – intensive care unit.

Outcome: drain-related meningitis

DRM was diagnosed by manual chart review using the adapted definition for healthcare-associated meningitis from the Centers for Disease Control's National Healthcare Safety Network[14,18,19]. In short, patients must either have microbiological growth from CSF cultures, or a combination of clinical signs and anomalities in CSF consistent with meningitis and administration of antibiotic therapy directed at DRM. Importantly, cultures with common skin contaminants in the absence of clinical signs reflective of meningitis are not considered cases of DRM and infections must occur within seven days of drain removal to be attributed to the EVD or ELD.

In hospitals B and C the definition was applied by infection control professionals using traditional manual surveillance (with confirmation by a neurosurgeon or a second infection control professional). In hospital A, the initial chart review was applied by an infection control researcher and possible cases were discussed with a physician for final adjudication (**table 1**). Prior to study initiation, all members of the research team performing manual chart review attended a training session reviewing definitions for DRM, methods of data collection and possible pitfalls as to improve consistency and quality of surveillance.

Electronic data collection and surveillance systems

Data required by the automated surveillance systems were extracted from electronic microbiology, pharmacy and clinical chemistry databases, using existing clinical data warehouses wherever possible[22]. Device utilization and drain characteristics were derived from traditional surveillance results. As in previous studies, a surveillance window was defined for each patient as the day of first drain placement until seven days after removal of the last drain or discharge and the value most reflective of (possible) DRM observed during this window was used in the surveillance systems (e.g., highest CSF leukocyte count). If a specified laboratory test was not performed throughout the surveillance window, it was classified as missing. Data were handled and processed exactly as in model development. Importantly, only measurements performed during routine patient care were included and no changes to existing diagnostic or therapeutic protocols were made for the purpose of this study.

The automated surveillance system based on the multivariable regression model has previously been described and is summarized in the appendix (**Figure S1**, **Table S1**)[14,23]. In brief, the multivariable logistic regression model uses data on microbiological culture results, antibiotic dispensing and clinical chemistry results combined with drain

characteristics to estimate a predicted probability of DRM for each patient. Missing values are handled by multiple imputation[24,25]. The regression model returns predicted probabilities of DRM (P(DRM)) that can subsequently be used to stratify patients as high or low risk of DRM based on a pre-specified threshold or to estimate the total number of infections without the requirement for manual review[14].

An alternative, less complicated, automated surveillance strategy based on a classification algorithm was also implemented. In such a system patients were flagged as high or low risk simply based on the presence or absence of indicators of infection (positive culture, antibiotic exposure) instead of incorporating them in a, weighted, multivariable regression formula[2]. Missing observations are classified as negative in this approach.

Analyses

After data linkage and verification of integrity, descriptive statistics were generated. Readmissions more than 30 days after discharge were considered independent drainage episodes. As with model development, missing data patterns were assessed (data not shown) and multiple imputation was performed, stratified by hospital. The multivariable regression formula was applied to each imputed dataset and the predicted probabilities were averaged (**table S1**)[14]. To assess the performance of the multivariable regression approach, the area under the ROC curve was used to assess model discrimination and sensitivity, specificity, predictive values and workload reduction were calculated for a range of predicted probability thresholds, including a previously specified threshold (P(DRM) = 0.15)[2]. Calibration plots were examined and the overall number of infections within a specific group of patients was estimated without any manual review by summing all the predicted probabilities within that group (calibration-in-the-large)[26]. In addition, the incremental benefit of the multivariable regression approach as compared to the less complicated classification algorithm was investigated for its ability to detect DRM cases identified by traditional manual surveillance by examining sensitivity, positive predictive value and workload reduction. All analyses were performed using SPSS Statistics 20 (IBM, Armonk NY) and R version 3.0.1 (www.r-project.org).

RESULTS

During the study period, 419 patient admissions with one or more external ventricular and/or lumbar drains were recorded. Of these, 12 patients were children, 19 died within 24 hours of drain placement, 18 had pre-existing meningitis and four were transferred with a drain in situ, leaving 366 drainage episodes in 354 unique patients eligible for analysis.

Table 2 | Population characteristics by hospital.

	Total	Hospital A	Hospital B	Hospital C
Median (IQR), n (%)	**n = 366**	**n = 62**	**n = 200**	**n = 104**
Age	59 (48-68.2)	64 (53-73)	58 (45-67)	59 (49-67)
Female	203 (55.5)	36 (58.1)	110 (55.0)	57 (54.8)
Indication for drain placement				
- Intracranial hemorrhage	166 (45.4)	47 (75.8)	66 (33.0)	53 (51.1)
- Cerebral infarction	3 (0.8)	1 (1.6)	1 (0.5)	1 (1.0)
- CSF leakage	91 (24.9)	4 (6.5)	73 (36.5)	14 (13.5)
- Per-operative prophylaxis	69 (18.9)	5 (8.1)	41 (20.5)	23 (22.1)
- Trauma	4 (1.1)	1 (1.6)	2 (1.0)	1 (1.0)
- Tumor	5 (1.4)	1 (1.6)	0	4 (3.8)
- Other	28 (7.7)	3 (4.8)	17 (8.5)	8 (7.7)
Admitted to ICU	171 (46.7)	42 (67.7)	71 (35.5)	58 (55.8)
Prior neurosurgery (< 30 days)	135 (36.9)	8 (12.9)	97 (48.5)	30 (28.8)
One or more EVD placed[a]	185 (50.5)	54 (87.1)	79 (39.5)	52 (50.0)
Only ELD placed[a]	181 (49.5)	8 (12.9)	121 (60.5)	52 (50.0)
Total number of drains[a]	1 (1-1)	1 (1-2)	1 (1-1)	1 (1-1)
Total drain duration (days)	7 (5-11)	7 (5-12)	7 (5-11)	8 (4-11)
Blood leukocytes (max, $x10^9$/L)[a]	16.5 (13.0- 20.8)	17.1 (14.2-21.1)	15.9 (12.3-20.7)	16.8 (13.1-20.7)
C-reactive protein (mg/L)[a]	59 (23-124)	56 (24-104)	42 (12-105)	99 (38-177)
CSF leukocytes (x100/uL)[a]	1.9 (0.3-11.4)	8.0 (1.4-26.7)	1.4 (0.3-7.3)	1.8 (0.2-6.9)
CSF glucose	3.2 (2.2-4.1)	3.0 (2.1-3.5)	3.2 (2.1-4.2)	3.5 (2.9-4.5)
CSF total protein	1.2 (0.6)	1.8 (0.9-3.7)	0.8 (0.5-1.7)	1.2 (0.4-2.1)
Positive culture from CSF or drain[a]	49 (13.4)	18 (29.0)	19 (9.5)	13 (12.5)
Positive gram stain from CSF[a]	23 (6.3)	6 (9.7)	12 (6.0)	5 (4.8)
Any antibiotic initiated	245 (66.9)	52 (83.9)	124 (62.0)	69 (66.3)
Empiric antimicrobial therapy[a]	54 (14.8)	8 (12.9)	31 (15.5)	15 (14.4)
Number antimicrobial switches[a]	1 (0-2)	1.5 (0-3)	1 (0-1)	1 (0-2)
Deceased in hospital	73 (20.3)	22 (39.3)	35 (17.5)	16 (15.4)

[a] in regression model. Notes: Antibiotics do not include per-operative prophylaxis. For laboratory values, the value most indicative of meningitis measured throughout the drainage episode is presented. There were no missing data except blood leukocytes (15.6%), C-reactive protein (23.0%), CSF leukocytes glucose and protein (45.3%), CSF or drain culture (43.4%), Gram stain (45.6%); data are presented prior to multiple imputation. Values were available for all predictors in 46.4% of cases (83.3%, 37.0% and 43.3% for hospitals A, B and C respectively).

Abbreviations: CSF – cerebrospinal fluid, ELD – external lumbar drain, EVD – external ventricular drain, ICU – intensive care unit, IQR – interquartile range.

Electronic collection of data from EHRs was feasible for all predictors with the exception of antibiotic use in one hospital, where extraction required manual data handling. Clinical data warehouses in hospitals A and C greatly facilitated data extraction.

Table 2 shows patient characteristics by hospital as well as the distribution of the predictors included in the model. Indication for drain placement varied between hospitals; three quarters of drains were placed to treat hydrocephalus after intracranial hemorrhage in hospital A, whereas in hospital B drains were often placed per-operatively or as treatment for CSF leakage. Approximately half of drainage episodes included one or more EVDs, and 181 only had ELDs (12.9% in hospital A, 60.5% in hospital B and 50% in hospital C). In just over 50% of drainage episodes, there were missing values for one of or more of the laboratory predictors used by the model, although this varied considerably between hospitals. Hospital A performed daily CSF sampling, hence the lower rates of missing observations. The overall rate of DRM was 12.3/1000 drain days at risk (4.6 for ELD and 18.4 for EVD), varying from 9.3 to 22.5 between hospitals (**table 3**). In 11 out of 37 DRM cases (29.7%) no causative micro-organism was isolated.

The overall area under the ROC curve after application of the regression model was 0.969 (95% confidence interval 0.946-0.991); ranging from 0.910 to 1.00 by hospital (**table 4**). Applying a cut-off of 0.15 to the predicted probabilities resulted in 97.3% sensitivity for DRM, 52.2% positive predictive value and reduced the number of charts that require manual review by 81%. Using higher thresholds, for example 0.20 or 0.25, has higher

Table 3 | Rates of drain-related meningitis, stratified by hospital.

	Total	Hospital A	Hospital B	Hospital C
	n = 366	n = 62	n = 200	n = 104
n DRM	37	12	17	8
Total drain days at risk	3010	533	1616	861
- Rate/1000 drain days at risk	12.3	22.5	10.5	9.3
- Rate/1000 drain days at risk (EVD)	18.4	25.3	15.5	15.7
- Rate/1000 drain days at risk (ELD)	4.6	0	6.6	0
Causative micro-organisms				
- None cultured	11	4	4	3
- Coagulase-negative staphylococci	16	7	6	3
- *Staphylococcus aureus*	1	0	1	0
- Enterobacteriaceae	4	0	3	1
- Enterococcus species	3	0	2	1
- Streptococcus species	1	1	0	0
- Bacillus species	1	0	1	0

Abbreviations: DRM – drain-related meningitis, ELD – external lumbar drain, EVD – external ventricular drain.

Table 4 | Performance characteristics of the multivariable model, stratified by hospital, with 95% confidence intervals.

	Total	Hospital A	Hospital B	Hospital C
	n = 366	**n = 62**	**n = 200**	**n = 104**
Area under ROC curve	0.969	0.910	0.969	1.000
(95% CI)	(0.946-0.991)	(0.806-1.000)	(0.947-0.992)	(1.000-1.000)
P(DRM) = 0.15				
- Sensitivity	97.3 (85.8-99.9)	91.7 (61.5-99.8)	100 (80.5-100)	100 (63.1-100)
- PPV	52.2 (39.8-64.4)	52.4 (29.8-74.3)	50.0 (32.4-67.6)	57.1 (28.9-82.3)
- % charts to review	18.9	33.9	17.0	13.5
P(DRM) = 0.20				
- Sensitivity	97.3 (85.8-99.9)	91.7 (61.5-99.8)	100 (80.5-100)	100 (63.1-100)
- PPV	59.0 (45.7-71.5)	61.1 (35.8-82.7)	54.8 (36.0-72.7)	72.7 (39.0-94.0)
- % charts to review	16.7	29.0	15.5	10.6
P(DRM) = 0.25				
- Sensitivity	81.1 (64.9-92.0)	75.0 (42.8-94.5)	76.5 (50.1-93.2)	100 (63.1-100)
- PPV	62.5 (45.0-72.4)	60.0 (32.3-83.7)	59.1 (36.4-79.3)	72.7 (39.0-94.0)
- % charts to review	13.1	24.2	11.0	10.6

Notes: In model development, the P(DRM) = 0.15 cut-off resulted in 98.1% sensitivity, 87.6% specificity and PPV of 52.5% with 26% of charts requiring manual review. Confidence intervals are calculated by the exact binomial method.

Abbreviations: CI – confidence interval, P(DRM) – predicted probability of drain-related meningitis, PPV – positive predictive value, ROC – receiver operating characteristics.

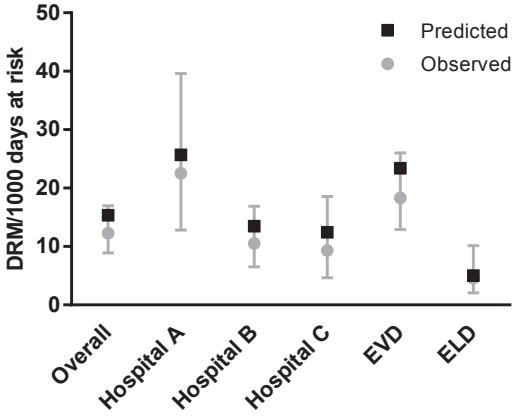


Figure 1 | Observed versus predicted rates of drain-related meningitis, stratified by hospital and drain type.

The predicted number of meningitis cases is derived by summing the predicted probabilities derived from the model within each group, without applying manual chart review. Gray bars represent 95% confidence intervals for the observed rates.Abbreviations: DRM – drain-related meningitis, ELD – external lumbar drain, EVD – external ventricular drain.

Table 5 | Performance of classification algorithms. Patients are classified as high-or low risk based on the simple presence or absence of makers of infection (instead of application of a regression model). Performance of the regression model is also presented for comparison.

	Model components	Hospital A			Hospital B			Hospital C		
		Sens (%)	PPV (%)	% rev	Sens (%)	PPV (%)	% rev	Sens (%)	PPV (%)	% rev
1	Culture or Gram stain	88.3	52.6	30.6	64.7	50.0	11.0	87.5	53.8	12.5
2	Culture or Gram stain, contaminant correction[a]	66.7	61.5	21.0	64.7	52.4	10.5	75.0	66.7	8.7
3	Any antimicrobial exposure (> 4 days)	100	27.9	69.4	100	19.3	44.0	100	23.5	32.7
4	Antimicrobial exposure (empiric regimen)	50.0	75.0	12.9	76.5	41.9	15.5	87.5	46.7	14.4
5	2 *OR* 4	91.7	64.7	27.4	100	41.5	20.5	100	42.1	18.3
6	Regression model (P(DRM) = 0.15)	91.7	52.4	33.9	100	50.5	17.0	100	57.1	13.5

[a] contaminant correction is as in the multivariable regression model (table S2 for details).

Notes: missing observations are classified as negative in the classification algorithms.

Abbreviations: PPV – positive predictive value, P(DRM) – predicted probability of drain-related meningitis, Sens – sensitivity, % rev – percentage of charts requiring manual review.

efficiency, albeit at the cost of lower sensitivity for the latter. As an exploratory analysis, reasons for misclassification of DRM were investigated (**table S2**). Interestingly, in hospital A, 5 of 10 false-positives were identified based on clinical chemistry results alone. **Figure 1** depicts observed and predicted rates of DRM when the model is applied without any manual review of charts. Examination of calibration plots showed adequate overall calibration, although stratification by hospital showed some miscalibration (**figure S2**).

Table 5 compares the performance of the multivariable regression approach to the alternative, less complex, automated classification algorithm with respect to sensitivity and efficiency of surveillance. For example, if all patients with some evidence of microbiological infection of the CSF or drain were to undergo manual chart review, sensitivity of DRM detection ranges from 64% to 88% by hospital, with between 11% and 31% of charts requiring manual review (option 1). If data on antibiotic exposure is used to correct for possible contamination with skin flora, sensitivity decreases but efficiency increases (option 2). The most optimal combination of contaminant-corrected culture results with data on empiric antibiotic therapy (option 5) resulted in 97.3% sensitivity, 46.8% positive predictive value and 21% of charts requiring review for all hospitals combined. In two of the three hospitals the multivariable regression model with the pre-specified threshold achieved higher efficiency at similar sensitivity than the classification algorithm (hospitals B and C).

DISCUSSION

This study demonstrates the external validity, both temporal and geographical, of a previously developed automated surveillance system for drain-related meningitis based on a multivariable regression model using routine care data extracted from EHRs. In comparison to a more straightforward classification algorithm, the regression model reached a more favorable sensitivity-to-efficiency balance in two of three hospitals. Despite important differences between hospitals with respect to underlying patient characteristics, diagnostic practices and treatment protocols, the regression model could be used to accurately and efficiently identify those patients requiring manual chart review with near-perfect sensitivity and achieved an average workload reduction of over 80%. Furthermore, estimation of overall DRM rates without the need for any manual chart review was fairly accurate. All cases of DRM in which no causative micro-organism was isolated were identified.

The (semi-)automated multivariable regression approach to surveillance for DRM had very similar performance characteristics to those observed during model development[14], thus providing evidence for its generalizability to other settings. Additional benefits of regression models compared to classification algorithms are the flexibility of adapting thresholds to specific situations and the possibility to directly estimate group-level infection rates by summing the predicted probabilities. Importantly, application of classification algorithm requires similar data extraction and processing as the regression model. The findings of this study support the notion that the use of multivariable regression models may be of incremental value in automated surveillance of HAI[2], Previous research has demonstrated important difference between hospitals in methods of HAI surveillance and application of definitions[11,27,28]. Implementation of automated surveillance and hence consistent identification of high-risk cases may in the future contribute to more standardized HAI surveillance and thus assist in efficiently generating reliable HAI rates[8].

Extraction from EHRs was possible for nearly all predictors although some effort had to be made to standardize data prior to model application and care must be taken to ensure comparability of electronic data sources across hospitals. Further expansion of clinical data warehouses, standardized documentation and electronic data capture are expected to facilitate model implementation in practice[29]. Importantly, in this study data on device use (denominators) were collected by manual surveillance as none of the hospitals currently have structured documentation of EVD and ELD usage in a format that is amenable to electronic data extraction. In analogy with other device-associated infections, these concerns will need to be addressed in order to make maximal benefit of automated surveillance[30,31].

As in model development, the use of routine care data was associated with missing predictor values, albeit that a larger fraction of patients had missing values, in particular in hospitals B and C. Multiple imputation was a feasible method of handling these missing observations in this context, and future work will need to identify the optimal method of applying the model in practice[32]. Interestingly, as a result of daily CSF sampling in hospital A, 83.3% of patients had no missing predictor information but model efficiency was lower compared to the other hospitals; this may in part be explained by the higher number of true infections and the increased likelihood of outlying measurements especially since a relatively large fraction of patients incorrectly flagged as high-risk were due to – perhaps incidental – anomalities in laboratory values.

In this study, both regular and antibiotic-impregnated EVDs were in use although no distinction was made between these in model application. Sensitivity analyses showed very similar incidence of DRM for these drain types and no difference in performance (data not shown). It was not the aim of this study to evaluate absolute rates of DRM, and differences in DRM rates observed between the three centers may be explained by (a combination of) variations in underlying risk, indications for drain placement, types of drains placed, clinical and diagnostic practices.

This study has several limitations. As with many studies on methods of HAI surveillance, ascertainment of infection status is not straightforward[33]. Not only are there many different definitions for DRM currently in use[15,19], the complex nature of the definition makes it vulnerable to error and subjective interpretation. We aimed to minimize this possible source of variability by providing all study personnel with a training session using prepared case vignettes. Notwithstanding these concerns, however, the purpose of the automated surveillance systems presented in this study is to increase efficiency of surveillance beyond what is currently available and in the absence of a perfect reference standard we have used the best-available option. Furthermore, this study did not include post-discharge surveillance although patients readmitted within a week of previous discharge were linked with the prior episode. As opposed to HAI that require a longer duration of follow-up, the consequences are expected to be minimal for DRM as meningitis should by definition occur within 7 days of device removal and most patients will be readmitted to the original hospital in case of worsening clinical status.

Future work needs to address concerns with electronic data collection for both denominators (device use) and predictors of DRM, ensure ongoing validation of performance - as with any method of surveillance - and assess whether (semi-)automated methods are truly more reliable than manual approaches. Moreover, depending on

the desired application of the automated surveillance approach, hospital-specific customization may be considered in order to maximally support within hospital surveillance efforts, though perhaps at the cost of losing comparability across institutions. Recent developments in fuzzy-logic and machine learning may also contribute to more detailed automated identification of HAI[34].

Automated surveillance systems for DRM could accurately and efficiently identify patients at high risk of having developed DRM. The multivariable regression approach had a higher efficiency than conventional automated classification algorithms in two of three hospitals and implementation can decrease the burden for manual review by over 80% at almost perfect sensitivity. This approach may be of incremental value when developing (semi-) automated systems for HAI surveillance. Future work is required to enable electronic data collection on device use and provide ongoing validation of automated surveillance approaches.

Acknowledgments

We would like to acknowledge the following people for their invaluable contributions in collecting the data: F. Bruna, E. N. Koster-den Dulk, E. Kelleter-Grasma, A. Kleijweg, Y. Groenleer, H.E.M. Blok, S.M. van Dijk, L. Stijvers.

Funding

This work was supported by an Alexander Suermann stipend provided by the UMC Utrecht [MvM]. Development of electronic data extraction tools in Medisch Centrum Haaglanden was made possible by a grant from the MCH Research Fund.

Conflict of Interest

All authors report no conflict of interest relevant to this article.

SUPPLEMENTARY DATA

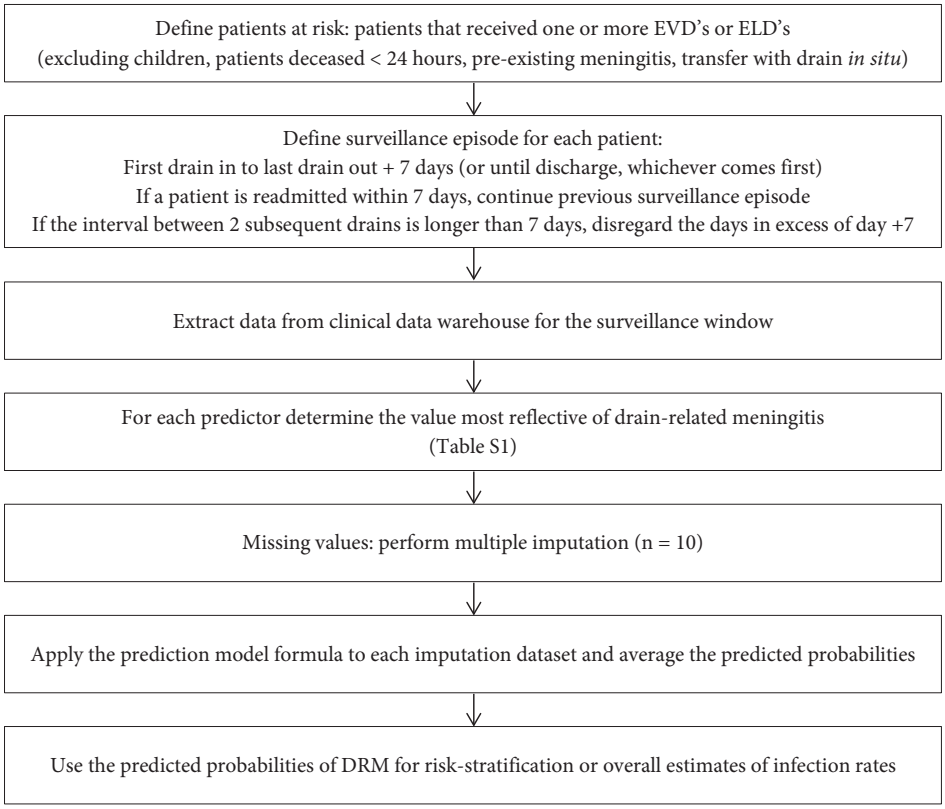


Figure S1 | Data processing flowchart for the multivariable regression approach.

Notes: Variables in multiple imputation model in addition to the predictors in the multivariable model and outcome were the following: age, sex, Indication for drainage, ICU admission status, total drain duration, hemoglobin levels, platelet count, neutrophil count (blood), Ln(CSF erythrocytes), glucose (CSF), protein (CSF), total number of antibiotic and corticosteroid prescriptions, antibiotic exposure data (total number, percentage of days exposed, any exposure to regimens > 4 days, average duration of antibiotic regimens, number of days exposed to empiric antimicrobial therapy, any exposure to flucloxacillin, ceftazidim, vancomycine, meropenem, ceftriaxone), any corticosteroid exposure**,** percentage of days exposed to corticosteroids, total number of days exposed to corticosteroids.

Abbreviations: DRM – drain related meningitis, ELD – external lumbar drain, EVD – external ventricular drain.

Table S1 | Predictor specifications and model parameters applied.

Predictor	Details	coefficient
Drain type	If patients received only ELD(s): code as 0 One or more EVD(s), including antibiotic-coated: code as 1.	1.486
Number of drains	Total number of drain exchanges. If two drains are placed simultaneously, code as 2 drains	0.523
C-reactive protein	Highest value observed during surveillance window (in mg/L) For model application: divide by 10.	-0.077
Blood leukocytes	Highest value observed during surveillance window ($x10^6$/L)	0.082
CSF leukocytes	Highest value observed during surveillance window (x100/uL) For imputation and model application: Natural logarithm	0.202
Positive culture CSF/ drain or Gram stain CSF (corrected for contamination)	Any microbial growth from CSF or drain tip, or a Gram stain from CSF showing bacteria. If skin-contaminants are isolated, and no antibiotic therapy lasting more than 4 days is initiated from day -1 to day + 3 surrounding the culture, it is classified as negative.	2.495
Number of antibiotic switches	Total number of new antibiotics prescribed during the surveillance window. Dosage change or changes in formulation of the same generic compound do not qualify as a new prescription. Suspension used for selective decontamination of the digestive tract and other topical preparations are excluded.	0.203
Empiric therapy	Defined by local protocol for empiric treatment of drain-related meningitis (usually combination therapy targeting skin flora and gram-negatives, see main article for details). Coded as yes (1) if patients were exposed to one or more days of empiric treatment, else coded as no (0).	1.803

Intercept:-6.615. Hence the following formula is applied to each patient to calculate the probability of DRM.

$$P(DRM) = \frac{1}{1 + e^{-LP}}$$

$$\begin{aligned} LP = {} & -6.615 + 1.486 * \text{drain type} + 0.523 * \text{Number of drains} - 0.077 * \text{C reactive protein} \\ & + 0.082 * \text{blood leukocytes} + 0.202 * \text{CSF leukocytes} + 2.495 \\ & * \text{CSF culture Gram stain} + 0.203 * \text{Nr antibiotic switch} + 1.803 \\ & * \text{Empiric antibiotic therapy}. \end{aligned}$$

Abbreviations: CSF – cerebrospinal fluid, ELD – external lumbar drain, EVD – external ventricular drain.

Table S2 | Reasons for discrepancy at the P(DRM) = 0.15 threshold, stratified by hospital.

	Total	A	B	C
	n= 366	n = 62	n = 200	n = 104
False-negatives	1	1	0	0
- Incorrect contaminant correction by data processing	1	1		
False positives	33	10	17	6
- Possible infection, but rejected	13	3	7	3
- Empiric antibiotics + culture, disjoint in time	3	1	2	
- Empiric antibiotics only (no other signs)	7		5	1
- Isolated positive culture (observed)	3		1	2
- Isolated positive culture (due to imputation)	1	1		
- Combination of laboratory tests, no empiric antibiotics	7	5	2	

Abbreviations: P(DRM) – predicted probability of drain-related meningitis, EVD – external ventricular drain.

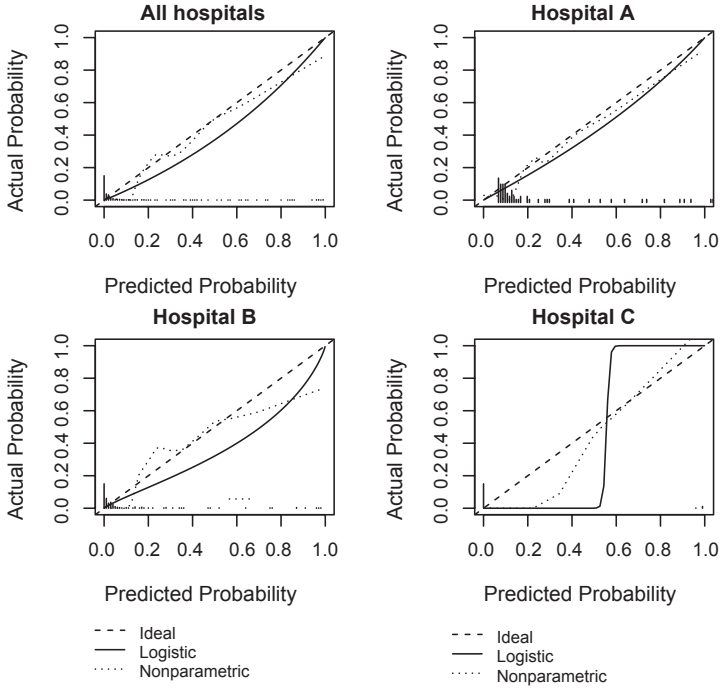


Figure S2 | Calibration plots for the multivariable prediction model, for all hospitals combined and stratified by hospital.

REFERENCE LIST

1. Freeman R, Moore LS, Garcia AL, Charlett A, Holmes A. Advances in electronic surveillance for healthcare-associated infections in the 21st Century: a systematic review. *J Hosp Infect.* **2013**;84:106-119.
2. van Mourik MS, Troelstra A, van Solinge WW, Moons KG, Bonten MJ. Automated surveillance for healthcare-associated infections: opportunities for improvement. *Clin Infect Dis.* **2013**;57:85-93.
3. Trick WE. Decision making during healthcare-associated infection surveillance: a rationale for automation. *Clin Infect Dis.* **2013**;57:434-440.
4. Tokars JI, Richards C, Andrus M et al. The changing face of surveillance for health care-associated infections. *Clin Infect Dis.* **2004**;39:1347-1352.
5. Haley RW, Culver DH, White JW et al. The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals. *Am J Epidemiol.* **1985**;121:182-205.
6. Haustein T, Gastmeier P, Holmes A et al. Use of benchmarking and public reporting for infection control in four high-income countries. *Lancet Infect Dis.* **2011**;11:471-481.
7. Geubbels EL, Nagelkerke NJ, Mintjes-De Groot AJ, Vandenbroucke-Grauls CM, Grobbee DE, De Boer AS. Reduced risk of surgical site infections through surveillance in a network. *Int J Qual Health Care.* **2006**;18:127-133.
8. Lin MY, Hota B, Khan YM et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. *JAMA.* **2010**;304:2035-2041.
9. Klompas M. Interobserver variability in ventilator-associated pneumonia surveillance. *Am J Infect Control.* **2010**;38:237-239.
10. Mayer J, Greene T, Howell J et al. Agreement in Classifying Bloodstream Infections Among Multiple Reviewers Conducting Surveillance. *Clin Infect Dis.* **2012**;55:364-370.
11. Klompas M. Eight initiatives that misleadingly lower ventilator-associated pneumonia rates. *Am J Infect Control.* **2012**;40:408-410.
12. Woeltje KF. Moving into the future: electronic surveillance for healthcare-associated infections. *J Hosp Infect.* **2013**;84:103-105.
13. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. *Clin Infect Dis.* **2009**;48:1268-1275.
14. van Mourik MS, Moons KG, van Solinge WW et al. Automated Detection of Healthcare Associated Infections: External Validation and Updating of a Model for Surveillance of Drain-Related Meningitis. *PLoS ONE.* **2012**;7:e51509.
15. Lozier AP, Sciacca RR, Romagnoli MF, Connolly ES, Jr. Ventriculostomy-related infections: a critical review of the literature. *Neurosurgery.* **2002**;51:170-181.
16. Scheithauer S, Burgel U, Bickenbach J et al. External ventricular and lumbar drainage-associated meningoventriculitis: prospective analysis of time-dependent infection rates and risk factor analysis. *Infection.* **2010**;38:205-209.
17. Lyke KE, Obasanjo OO, Williams MA, O'Brien M, Chotani R, Perl TM. Ventriculitis complicating use of intraventricular catheters in adult neurosurgical patients. *Clin Infect Dis.* **2001**;33:2028-2033.
18. Leverstein-Van Hall MA, Hopmans TEM, Van Der Sprenkel JWB et al. A bundle approach to reduce the incidence of external ventricular and lumbar drain-related infections: Clinical article. *J Neurosurg.* **2010**;112:345-353.
19. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control.* **2008**;36:309-332.
20. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* **2009**;338:b605.
21. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* **1999**;130:515-524.
22. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med.* **2007**;45:13-19.

23. van Mourik MS, Groenwold RH, Berkelbach van der Sprenkel JW, van Solinge WW, Troelstra A, Bonten MJ. Automated detection of external ventricular and lumbar drain-related meningitis using laboratory and microbiology results and medication data. *PLoS One.* **2011**;6:e22846.
24. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* **2006**;59:1087-1091.
25. Sterne JA, White IR, Carlin JB et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* **2009**;338:b2393.
26. Steyerberg EW. *Clinical Prediction Models: A practical approach to development, validation, and updating.* New York: Springer; **2009**.
27. Keller SC, Linkin DR, Fishman NO, Lautenbach E. Variations in identification of healthcare-associated infections. *Infect Control Hosp Epidemiol.* **2013**;34:678-686.
28. Dixon-Woods M, Leslie M, Bion J, Tarrant C. What counts? An ethnographic study of infection data reported to a patient safety program. *Milbank Q.* **2012**;90:548-591.
29. Trick WE. Building a data warehouse for infection control. *Am J Infect Control.* **2008**;36:S75-81.
30. Tejedor SC, Garrett G, Jacob JT et al. Electronic documentation of central venous catheter-days: validation is essential. *Infect Control Hosp Epidemiol.* **2013**;34:900-907.
31. Wright MO, Fisher A, John M, Reynolds K, Peterson LR, Robicsek A. The electronic medical record as a tool for infection surveillance: successful automation of device-days. *Am J Infect Control.* **2009**;37:364-370.
32. Janssen KJ, Vergouwe Y, Donders AR et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem.* **2009**;55:994-1001.
33. Lin MY, Bonten MJ. The Dilemma of Assessment Bias in Infection Control Research. *Clin Infect Dis.* **2012**;54:1342-1347.
34. De Bruin JS, Blacky A, Koller W, Adlassnig KP. Validation of fuzzy sets in an automated detection system for intensive-care-unit-acquired central-venous-catheter-related infections. *Stud Health Technol Inform.* **2013**;192:215-218.

6

Electronic implementation of a novel surveillance paradigm for ventilator-associated events: feasibility and validation

Maaike S.M. van Mourik[a*], Peter M.C. Klein Klouwenberg[a,b,c*], David S.Y. Ong[a,b,c], Janneke Horn[d], Marcus J. Schultz[d], Olaf L. Cremer[b], Marc J.M. Bonten[a,c]
*Contributed equally to this work

[a] *Department of Medical Microbiology, University Medical Center Utrecht*
[b] *Department of Intensive Care Medicine, University Medical Center Utrecht*
[c] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht*
[d] *Department of Intensive Care Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands*
This study was performed on behalf of the MARS consortium

American journal of Respiratory and Critical Care Medicine 2014 Vol 189(9): 947 - 955.
Reprinted with permission of the American Thoracic Society. Copyright © 2014 American Thoracic Society. Official Journal of the American Thoracic Society.

ABSTRACT

Rationale - Accurate surveillance of ventilator-associated pneumonia (VAP) is hampered by subjective diagnostic criteria. Recently, a novel surveillance paradigm for ventilator-associated events (VAE) was introduced.

Objectives - To determine the validity of surveillance using the new VAE algorithm.

Methods - Prospective cohort study in two Dutch academic medical centers (2011-2012). VAE surveillance was electronically implemented and included assessment of (infection-related) ventilator-associated conditions (VAC, IVAC) and VAP. Concordance with ongoing prospective VAP surveillance was assessed, along with clinical diagnoses underlying VAEs and associated mortality of all conditions. Consequences of minor differences in electronic VAE implementation were evaluated.

Measurements and Main Results - The study included 2080 patients with 2296 admissions. Incidences of VAC, IVAC, VAE-VAP and VAP according to prospective surveillance were 10.0, 4.2, 3.2 and 8.0 per 1000 ventilation days, respectively. The VAE algorithm detected at most 32% of the VAP patients identified by prospective surveillance. VAC signals were most often caused by volume overload and infections, but not necessarily VAP. Subdistribution hazards for mortality were 3.9 (95% CI 2.9 - 5.3) for VAC, 2.5 (1.5 - 4.1) for IVAC, 2.0 (1.1 - 3.6) for VAE-VAP and 7.2 (5.1 - 10.3) for VAP identified by prospective surveillance. In sensitivity analyses, mortality estimates varied considerably following minor differences in electronic algorithm implementation.

Conclusions - Concordance between the novel VAE algorithm and VAP was poor. Incidence and associated mortality of VAE were susceptible to small differences in electronic implementation. More studies are needed to characterize the clinical entities underlying VAE and ensure comparability of rates from different institutions.

INTRODUCTION

Healthcare-associated infections add a considerable burden to medical care. Monitoring and prevention of these infections have become increasingly important, along with a rise in (mandatory) public reporting[1-3]. Surveillance networks such as the National Healthcare Safety Network (NHSN) allow for benchmarking among hospitals, and feedback of infection rates offers guidance for improvement programs[4,5]. Ventilator-associated pneumonia (VAP) has been associated with increased mortality and length of stay. Therefore it is one of the major infections targeted by surveillance programs[6]. However, establishing the diagnosis of VAP is challenging and concerns have been voiced regarding the reliability of VAP surveillance and its use as a tool for hospital benchmarking[7-9]. VAP case definitions are complex, labor-intensive and leave room for subjective interpretation[10]. Low interrater reliability and poor correlation with histopathology have also been described[11,12]. In consequence, variations in implementation of VAP surveillance across hospitals affect the reported VAP rates and preclude valid interhospital comparisons[9,11,13]. Furthermore, studies of healthcare-associated infection surveillance and prevention are vulnerable to assessment bias[14].

These limitations of VAP surveillance have led to the development and implementation by the NHSN of a new surveillance paradigm for ventilated patients that aims to assess ventilator-associated events (VAEs)[15-17]. This algorithm identifies ventilator-associated conditions (VACs) and infection-related ventilator-associated conditions (IVACs) as entities for public reporting, using objective case definitions based on ventilator settings, antimicrobial prescriptions, temperature and leukocyte counts that are amenable to automated implementation. The algorithm further defines possible and probable VAP for within-hospital quality monitoring (**Table 1**).

Table 1 | Overview of conditions evaluated in this study.

Entity	Abbreviation	Brief definition
Ventilator-associated event	VAE	Includes VAC, IVAC, possible VAP and probable VAP.
Ventilator-associated condition	VAC	New, sustained, respiratory deterioration after a two-day baseline period of stability or improvement on mechanical ventilation.
Infection-related ventilator-associated condition	IVAC	VAC with evidence of infection (new antibiotics and inflammatory signs).
Ventilator-associated pneumonia (VAE)	VAE-VAP	IVAC with microbiological evidence of pneumonia (classified as possible or probable).
Ventilator-associated pneumonia	PROSP-VAP	Evidence of VAP by prospective surveillance definition (classified as definite, probable or possible VAP). Requires combination of clinical signs, radiographic and microbiological evidence.

This study aimed to assess the feasibility and validity of surveillance based on electronic implementation of the newly introduced VAE algorithm in a multicenter setting in The Netherlands. For this purpose, causes of VAC and IVAC signals were assessed in order to evaluate the face validity of VAE, VAE results were compared to ongoing prospective VAP registration, and mortality estimates were calculated. Sensitivity analyses were performed to evaluate several adaptions of the VAE algorithm and assess the robustness of several approaches to electronic implementation. Some results of this study have been previously reported in the form of abstracts[18,19].

METHODS

Study design and population

This cohort study was incorporated in the ongoing MARS (Molecular Diagnosis and Risk Stratification of Sepsis) project in the mixed intensive care units (ICUs) of two tertiary referral centers in The Netherlands[20]. The Institutional Review Board approved an opt-out consent method (IRB number 10-056C). For the current study, we analyzed all adult patients who had received two or more consecutive days of mechanical ventilation (MV) between January 1st 2011 and July 1st 2012. Patients with do-not-resuscitate orders and patients ventilated in the prone position were included in the analysis. Patients on rescue MV (high-frequency ventilation, extracorporeal membrane oxygenation) were included, but the days on rescue ventilation were excluded from the analysis.

Implementation of the ventilator-associated events algorithm

The main analysis was performed using the VAE algorithm implemented as specified by the current NHSN protocol using electronically extracted data (minute-to-minute ventilator settings, antibiotic use, microbiology data and clinical characteristics)[16]. Ventilator-associated conditions are defined by either a greater than 3-cm H_2O increase in daily minimum positive end-expiratory pressure (PEEP) or a greater than 20% increase in fraction of inspired oxygen (FiO_2). See supplementary text for details.

In sensitivity analyses, we assessed two sets of alternative implementations of the original algorithm. The first implementation aimed to identify more representative episodes of respiratory deterioration by varying the required increases in levels of PEEP and FiO_2 to more than 5 cm H_2O and more than10%, respectively. In addition, the original algorithm does not incorporate the time patients are in spontaneous breathing trials when being weaned off the ventilator or when they are on nighttime ventilation. As this ignores the best condition of the patient for that day, a sensitivity analysis was performed that classifies respiratory deterioration leading to discontinuation of weaning trials as a VAC.

This concept was implemented by setting the PEEP to room air conditions (PEEP=0 if the patient was in a spontaneous weaning trial for more than 2.5 hours (10%) of the day (intermittent MV rule, **Figure 1**).

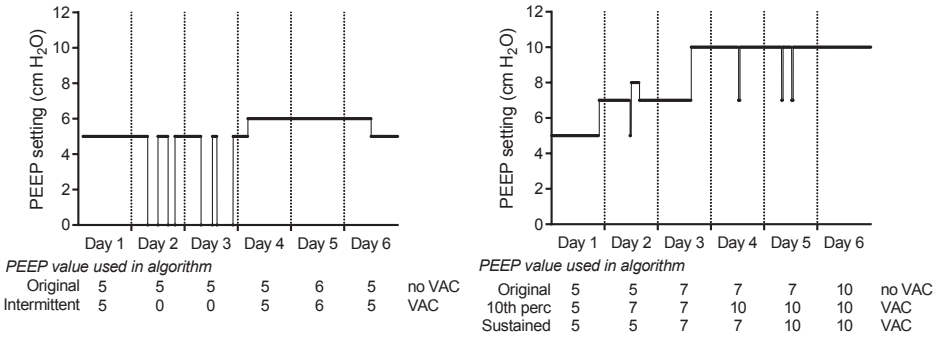


Figure 1 | Left panel – hypothetical example of a patient on intermittent ventilation under the original algorithm and the intermittent ventilation sensitivity analysis.

If mechanical ventilation was interrupted for more than 2.5 hours (i.e. 10% of the day) then daily minimum values were set to room air conditions (PEEP=0 and FiO_2=21%). Right panel – example of artifact filtering by both the 10th percentile (lowest 10 percent of daily measurements excluded) and the sustained settings rule (settings must be maintained for at least 60 consecutive minutes to qualify). The duration of the dip in PEEP on day 4 was 80 minutes; the two dips on day 5 were both 30 minutes.

Abbreviations: 10th perc – 10th percentile, PEEP – positive end-expiratory pressure, VAC – ventilator-associated condition.

The second analysis aimed to evaluate the robustness of electronic implementation of the VAE algorithm, as the current guidelines do not precisely specify the requirements for electronic data capture. As opposed to the original implementation, which involves the use of minute-to-minute measurements, we evaluated the effect of sampling frequency by using hourly, manually validated measurements only, which may better reflect common availability of data in electronic form. Second, to assess the effect of possible artifacts in the data, we excluded outliers from the minute-to-minute measurements by selecting the minimum daily ventilator settings either after application of a 10th percentile cut-off (thus excluding the lowest 10% of measurements, **Figure 1**) – 10th percentile rule – or by selecting the lowest setting maintained for at least one consecutive hour (sustained settings rule).

Reference standard

Patients were assessed daily for the development of VAP by dedicated well-trained observers with ongoing evaluation of inter-rater reliability[20]; this was considered the reference standard (**Table S1**). VAP was classified as definite, probable or possible, and assessment was fully independent of the VAE algorithm.

Alternative diagnoses

In order to assess what conditions may lead to a VAC signal, all patients flagged by the VAC algorithm (10th percentile rule) from one hospital were manually reviewed by two independent physicians (PK, MvM), with consensus through discussion with a third reviewer if necessary. The diagnoses that we explicitly considered are defined in **Table 2**. Reviewers were blinded to the patients' IVAC and VAP status.

Table 2 | Clinical conditions occurring in the five-day window surrounding VAC and IVAC events (hospital A).

Diagnosis	Identified by	Frequency identified	
		VAC (%) n = 81	IVAC (%) n = 31
Pulmonary conditions			
Respiratory tract infection*	Imaging, cultures, initiation of antibiotics, clinical documentation	23 (28)	14 (45)
Atelectasis/sputum plug	Bronchoscopy, imaging, explicit clinical documentation	12 (15)	5 (19)
Pneumothorax	Chest tube placement	2 (2)	--
Pulmonary embolus	Therapeutic anticoagulants	--	--
Pleural effusion	Pleural drainage	10 (12)	3 (10)
Aspiration	Explicit clinical documentation	5 (6)	1 (3)
Extra-pulmonary infection			
New onset of SIRS/sepsis	Clinical documentation, microbiology, initiation of antimicrobials	9 (11)	5 (16)
Cardiac/circulatory			
Volume overload	Initiation of diuretics	23 (28)	12 (39)
Heart failure	Inotropy or afterload reduction	6 (7)	4 (13)
Other			
Abdominal distension	Laparotomy, ascites drainage	9 (11)	4 (13)
Acute neurological event	Imaging & clinical findings	10 (12)	3 (10)
No reason for VAC identified		10 (12)	2 (6)

Notes: VAC was identified using the 10th percentile rule. Patients may have more than one alternative diagnosis. Sources of extra-pulmonary infection were: bloodstream (2), abdominal (5), mediastinal (1), and muscoskeletal (1). Acute neurological events included cerebrovascular accidents (5), cases of increased intracranial pressure (1), encephalopathy (3), and meningitis (1).

*Includes ventilator-associated pneumonia and (ongoing) other respiratory tract infections such as worsening pre-existing pneumonia.

Abbreviations: IVAC – infection related ventilator-associated condition, SIRS – Systemic Inflammatory Response Syndrome, VAC – ventilator-associated condition.

Statistical analyses

All ICU admissions were included to assess population characteristics and concordance between surveillance methods. Concordance between prospective VAP surveillance, stratified by VAP likelihood, and the VAE algorithm was assessed, both at the ICU admission level and by using a window of ± 2 days surrounding VAC, IVAC or VAE-VAP. This is analogous to the original algorithm that defines IVAC and VAE-VAP based on a five-day window around VAC.

We assessed the effects of VAEs and VAP on ICU mortality using competing-risk analyses. The direct effects of VAEs and VAP on outcome were estimated by the cause-specific hazard ratios (CSHRs) for each event (ICU discharge or ICU death). To evaluate the direct effect of the events on death, taking the competing event (discharge) into account, we calculated the subdistribution hazard ratio (SHR)[21]. All analyses were adjusted for age, gender, Acute Physiology and Chronic Health Evaluation (APACHE) IV score, admission type (surgical vs. medical) and hospital. VAC, IVAC, VAE-VAP and prospective VAP were included as time-dependent variables[22,23]. All analyses were performed using SAS 9.2 (Cary, NC), R version 2.14 (www.r-project.org) and SPSS 20 (IBM Software, Armonk NY).

RESULTS

During the study period, 3473 patients were admitted to the ICU, of whom 2080 patients (2296 admissions) were ventilated for at least two consecutive calendar days. Median age was 62 years and 44% were surgical patients. Overall ICU mortality was 21% (**Table 3**).

Using the original algorithm, 158 VACs were detected in 152 patients (10.0/1000 days of MV). Most events (n=149) were triggered by increasing PEEP settings, five by increased FiO_2 levels, and four by both. There were 66 IVACs in 65 patients (4.2/1000 MV days). All IVAC episodes that fulfilled the antibiotic exposure criteria also met the temperature and/or white blood cell count definition. There were 51 episodes of possible or probable VAP according to the VAE algorithm (3.2/1000 MV days) in 50 patients and 127 episodes of possible, probable or definite VAP were identified by prospective surveillance in 115 patients (8.0/1000 MV days) (**Table 3**). A large fraction of VAC (and IVAC) events occurred on the third or fourth day of mechanical ventilation (46%). Of the 2296 admissions, 108 did not achieve a baseline period of stability; of these, 60 deceased within 4 days of onset of ventilation. The incidence of VAE was higher in patients receiving seven or more days of mechanical ventilation compared to patients with less than seven days of mechanical ventilation (VAC 3.6 versus 12.9 per 1000 days of MV, IVAC 0.6 vs. 5.8 per 1000 days of MV). Rates of VAC, IVAC, VAE-VAP and prospectively monitored VAP were comparable between both ICUs (**Table S2**).

Table 3 | Patient characteristics and incidence of ventilator-associated events and ventilator-associated pneumonia.

	All patients	VAC positive	IVAC positive	VAE-VAP	PROSP-VAP
N (%) or median (IQR)	**n = 2296**	**n = 152**	**n = 65**	**n=50**	**n = 115**
Number of events	--	158	66	51	127
Incidence (/1000 MV)	--	10.0	4.2	3.2	8.0
Age	62 (50 – 72)	60 (52 – 69)	57 (49 – 66)	57 (48 – 62)	62 (49 – 71)
Male	1406 (61)	102 (67)	41 (63)	32 (64)	76 (66)
Comorbidities					
Cerebrovascular disease	164 (7)	9 (6)	1 (2)	0	9 (8)
Congestive heart failure	171 (7)	14 (9)	4 (6)	3 (6)	6 (5)
COPD	247 (11)	18 (12)	8 (12)	5 (10)	16 (14)
Diabetes	372 (16)	26 (17)	9 (14)	8 (16)	8 (7)
Malignancy	334 (15)	31 (20)	12 (18)	9 (18)	16 (14)
Admission type					
Medical	1303 (57)	91 (60)	45 (69)	37 (74)	61 (53)
Surgical emergency	567 (25)	36 (24)	8 (12)	5 (10)	37 (32)
Surgical elective	426 (19)	25 (16)	12 (18)	8 (16)	17 (15)
Admission source					
Operating theatre	640 (27)	37 (24)	14 (21)	9 (18)	31 (27)
Emergency department	810 (35)	51 (34)	18 (28)	13 (26)	48 (42)
Other ward	647 (28)	47 (31)	25 (38)	22 (44)	23 (20)
Other ICU	199 (9)	17 (11)	8 (12)	6 (12)	13 (11)
Readmission	324 (14)	22 (14)	15 (23)	14 (28)	12 (10)
Primary specialty					
Cardiothoracic surgery	347 (15)	24 (16)	10 (15)	6 (12)	25 (22)
Neurology or neurosurgery	540 (24)	18 (12)	8 (12)	8 (16)	33 (29)
Other surgery	648 (28)	43 (28)	17 (26)	13 (26)	20 (17)
Internal medicine	707 (31)	64 (42)	29 (45)	22 (44)	36 (31)
Other, unknown	54 (2)	3 (2)	1 (2)	1 (2)	1 (1)
APACHE IV	75 (57 – 97)	82 (65 – 105)	79 (62 – 102)	84 (56 – 102)	77 (56 – 100)
Duration of MV	4 (2 – 8)	13 (8.5 – 27)	15 (12 – 19)	15 (10 – 29)	14 (10 – 28)
Length of ICU stay	5 (2 – 9)	14 (10 – 29)	19 (13 – 33)	18 (12 – 33)	15 (12 – 34)
Deceased in ICU	476 (21)	56 (37)	18 (28)	12 (24)	39 (34)

Notes: VAE defined according to original algorithm, PROSP-VAP is as defined by the MARS study (includes possible, probable or definite).

Abbreviations: APACHE – Acute Physiology and Chronic Health Evaluation, COPD – chronic obstructive pulmonary disease, ICU – intensive care unit, IVAC – infection related ventilator-associated condition, VAC – ventilator-associated condition, VAE – ventilator-associated event, (PROSP –) VAP – (prospective) ventilator-associated pneumonia.

Concordance of VAC, IVAC and VAP

The sensitivity of VAC for detection of (possible, probable or definite) VAP was 33% (95% confidence interval [CI] 25 – 42%), with a positive predictive value of 25% (95% CI 18 – 33%) (**Table 4**). For IVAC and VAE-VAP the sensitivities were 18% (95% CI 12 – 27%) and 17% (95% CI 10 – 25%), respectively, with positive predictive values of 32% (95% CI 21 – 45%) and 38% (25 – 53%). When restricting concordance to the VAE window (VAP by prospective surveillance must have occurred within the five-day window surrounding VAC), sensitivities for detecting VAP were 13, 9 and 6% for VAC, IVAC and VAE-VAP, respectively, with positive predictive values of 10, 18 and 16%. Thus most episodes of VAC and IVAC were not temporally associated with VAP.

When restricting the reference standard to probable and definite VAP, events of VAC, IVAC and VAE-VAP detected 44, 25, and 22% of cases of VAP, respectively. Positive predictive values were 9, 12 and 11% (**Table 4**). Among the 35 episodes of probable or definite prospective VAP (32 patients) there were 25 episodes that did not fulfill the criteria for VAC, either because there was no baseline period of stability (n=6) or no (sufficient) increase in ventilator settings (n=19).

In a retrospective analysis of underlying clinical conditions, pneumonia, either VAP or pre-existing pneumonia, appeared the most often observed cause of VAC (**Table 2**). The interrater agreement was moderate (kappa = 0.51). The positive predictive value of IVAC for all respiratory tract infections combined (VAP, hospital-acquired pneumonia, community-acquired pneumonia and other lower respiratory tract infection) increased

Table 4 | Concordance between prospective VAP surveillance and VAE events detected by the original VAE algorithm at the patient level.

		Prospective VAP					
		Definite	**Probable**	**Possible**	**Any**	**Absent**	**Total**
VAC	Present	1	13	24	38	114	152
	Absent	2	16	59	77	2067	2144
IVAC	Present	0	8	13	21	44	65
	Absent	3	21	70	94	2137	2231
VAE-VAP	Present	0	7	12	19	31	50
	Absent	3	22	71	96	2150	2246
Total		3	29	83	115	2181	2296

Notes: For VAE algorithm, VAP includes possible or probable for VAE and in prospective surveillance possible, probable and definite.
Abbreviations: IVAC – infection related ventilator-associated condition, VAC – ventilator-associated condition, VAE – ventilator-associated event, VAP – ventilator-associated pneumonia.

to 66% at the patient level and 36% when restricting concordance to the five-day window (data not shown). The sensitivity of IVAC for detection of these combined conditions, however, was 7 and 3%, respectively.

Adaptations in the algorithm

As the majority of VACs was identified by increases in PEEP settings, the effect of setting a higher trigger (>5 vs. >3 cm H_2O PEEP increase) was evaluated. This resulted in 51 VACs with a sensitivity of 7% and positive predictive value of 22% for the detection of VAP. Conversely, lowering the required increase of FiO_2 (>10 vs. >20%) resulted in 224 episodes of VAC in 213 patients, with a sensitivity of 43% and positive predictive value of 23% for the detection of VAP.

Using the algorithm with handling of intermittent weaning, 241 patients with 261 VACs were identified, with only 101 episodes concordant with the original algorithm. This adapted algorithm had higher sensitivity (50%) and similar positive predictive value (24%) with respect to concordance with prospective surveillance.

Reliability of electronic implementation

Applying the 10th percentile rule in the original algorithm did not change the overall incidence of VAC (158 events in 152 patients), but only 117 of the detected VAC episodes were identical. The 10th percentile modification yielded similar results as the original VAC algorithm with respect to concordance with VAP surveillance (data not shown). Applying the sustained settings rule resulted in 157 patients with VAC with a sensitivity of 34% and a positive predictive value of 25%. Using hourly (validated) measurements resulted in 152 episodes of VAC in 149 patients with a sensitivity of 30% and a positive predictive value of 23% for the detection of VAP at the patient level (13% sensitivity and 11% positive predictive value when examining episode-level concordance). Of the 152 episodes of VAC detected when using hourly (validated) measured, 113 identical episodes were detected by the 10th percentile and the sustained settings rule. Of the 158 episodes identified by the original algorithm, 104 (65%) were also identified by all other sensitivity analyses. Thus, although the total number of episodes identified was similar for all algorithms, there were differences between the algorithms in the types of episodes that were identified.

Association with mortality

All types of VAE and VAP were significantly associated with an increased hazard of ICU death when taking into accounting the competing events process (**Table 5**). The association was strongest for VAC (time-averaged SHR 3.9; 95% CI 2.9 – 5.3), and lower

for IVAC (SHR 2.5; 95% CI 1.5 – 4.1) and VAE-VAP (SHR 2.0; 95% 95% CI 1.1 – 3.6). VAP according to prospective surveillance had the highest subdistribution hazard ratio for death (7.2; 95% CI 5.1 – 10.3). Analysis of the cause-specific hazard ratios (CSHRs) revealed that VAC and VAP identified by prospective surveillance, but not the other VAEs, had a significant direct effect on the hazard of dying (**Table 5, Table S3**). In addition, all types of VAE resulted in a lower daily probability of being discharged from the ICU after the onset of VAE, exposing patients longer to a daily risk of dying in the ICU - thus the increased risk of dying in the ICU after VAE is mainly the result of prolonged stay in the ICU rather than the direct effect of VAE on mortality. There was no significant interaction between the type of ICU admission (medical or surgical) and the different conditions (VAC, IVAC, VAE-VAP, prospective VAP) with regards to their effect on estimated associated mortality although for VAE-VAP there was a trend for slightly higher associated mortality in surgical patients whereas the opposite was observed for VAP identified by prospective surveillance. Estimates of associated mortality for VAC identified using the various electronic implementations were SHR 6.3 (95% CI 4.8 – 8.4) for the 10th percentile rule; SHR 5.2 (95% CI 3.9 – 6.9) for the sustained settings rule; and SHR 6.3 (95% CI 4.7 – 8.5) for the hourly sampling scheme.

Table 5 | Multivariable subdistribution hazards model for ICU mortality to account for competing outcomes.

	VAC*	IVAC*	VAE-VAP*	PROSP VAP*
Crude mortality (%)	51/134 (38.1)	17/56 (30.4)	12/43 (27.9)	39/104 (37.5)
CSHR death (95% CI)	3.96 (2.43 -6.45)	0.98 (0.57-1.70)	1.11 (0.60-2.05)	2.00 (1.34-3.00)
CSHR discharge (95% CI)	0.38 (0.26-0.56)	0.47 (0.33-0.66)	0.56 (0.30-1.05)	0.45 (0.34-0.58)
SHR† (95% CI)	3.92 (2.88-5.34)	2.51 (1.52-4.12)	1.99 (1.11-3.58)	7.24 (5.09-10.3)§
Covariates (SHR)				
Age	1.00 (0.99-1.00)	1.00 (0.99-1.00)	1.00 (0.99-1.00)	1.00 (0.99-1.00)
Gender (male=ref)	1.01 (0.84-1.23)	1.02 (0.86-1.26)	1.04 (0.86-1.26)	1.04 (0.86-1.24)
APACHE IV	1.03 (1.02-1.03)	1.03 (1.02-1.03)	1.03 (1.03-1.03)	1.03 (1.02-1.03)
Hospital	1.32 (1.10-1.59)	1.34 (1.11-1.61)	1.34 (1.11-1.61)	1.34 (1.11-1.61)
Admission type (surgical = ref)	1.06 (0.87-1.31)	1.09 (0.86-1.30)	1.06 (0.86-1.13)	1.10 (0.90-1.35)

Notes: Definitions are based on the original algorithm. If a patient was admitted multiple times, one admission was randomly selected. In all models there was no significant interaction between hospital and event, or between admission type and event.

* No recovery from any of the events was assumed. † Time-averaged subdistribution hazard ratio due to the time-varying nature of the event (this means that the SHR may vary depending on the timing of the event, e.g., a VAC on day 3 may have a different SHR than a VAC on day 8). § The inclusion of only probable and definite VAP resulted in a similar SHR (7.45 (4.12-13.50)).

Abbreviations: APACHE IV – acute physiology and chronic health evaluation IV, CI – confidence interval, CSHR – cause-specific hazard ratio, IVAC – infection related ventilator-associated condition, SHR – subdistribution hazard ratio, VAC – ventilator-associated condition, VAE – ventilator-associated event, (PROSP –) VAP – (prospective) ventilator-associated pneumonia.

DISCUSSION

The development and implementation of a novel surveillance paradigm for ventilator-associated events exemplifies the ongoing efforts towards reliable surveillance of healthcare-associated infections, and in particular of ventilator-associated complications and pneumonia. In this study the VAC, IVAC and VAP events had poor concordance with prospective VAP surveillance, especially when restricting the analysis to the five-day window surrounding VAC events. Importantly however, the VAE algorithm aims at identifying a broader range of ventilator-associated complications and this is confirmed by findings from our retrospective review. Although a significant fraction of cases of VAC and IVAC could be attributed to pulmonary infections, albeit not limited to VAP, conditions such as volume overload, non-pulmonary infections and a variety of other causes were also commonly implicated. Occurrence of a VAE was associated with an increased risk of death in ICU, however not as strongly as the occurrence of VAP identified by prospective surveillance. Interestingly, both IVAC and VAP defined by the VAE algorithm were associated with lower likelihoods of ICU mortality than VAC, possibly indicating that other conditions than (pulmonary) infection were responsible for at least part of the associated mortality of VAC. Furthermore, although electronic implementation of the algorithm is feasible in two academic centers, subtle differences in the method of electronic implementation affect the events identified by the algorithm and their associated mortality. In the absence of detailed specifications, the algorithm was implemented using both detailed minute-to-minute data collection and a more practical hourly sampling scheme; interestingly both data sources identified different episodes of VAE in different patients and had similarly poor concordance with prospective VAP surveillance.

Several other studies have also found moderate concordance of VAC with VAP occurrence at the patient level and some association of VAC with ICU mortality and length of stay[15,24-26]. These studies have also shown that VAC reflects a broader scope of clinical conditions than VAP alone[24,27]. The higher incidence of VAC in these studies may have resulted from differences in implementation of the algorithm or differences between study populations. The present study is the first assessing the concordance of IVAC and VAP identified by the VAE algorithm, comparing it to a pre-existing prospective VAP surveillance and evaluating the reliability of electronic implementation.

Interpretations of findings and implications

The VAE algorithm has been implemented in the United States as a novel tool for surveillance and benchmarking of ventilator-associated complications, not limited to

pneumonia. However, interpretation of VAE rates remains difficult and their usefulness for quality improvement has not yet been established. Several criteria could be helpful in evaluating the validity of this novel surveillance entity. Above all, the outcome of interest should be clinically relevant and measure all aspects of what it is intended to assess. Although this cannot be formally tested, the retrospective analysis of alternative diagnoses in the present study shows that VAC and IVAC measure a diversity of conditions, some of which may not be associated with (quality of) ventilation practices. Moreover, the detection of VAP – one of the major target conditions for the VAE algorithm – was poor. The large number of VACs occurring on the third or fourth day of ventilation could be representative of ongoing clinical deterioration as opposed to insufficient quality of care. Furthermore, IVACs appeared to detect respiratory infections not related to mechanical ventilation. Ideally, a surveillance method should also identify differences between groups of patients that differ in their underlying risk of developing the event of interest. Because this study was not aimed at detecting differences between hospitals and no interventions for VAP were implemented during the period of study, this could not be evaluated. Furthermore, a key aspect that remains to be assessed is the preventability of conditions identified by the VAE algorithm and their effect on patient-centered outcomes. Intervention trials evaluating quality improvement programs targeted at VAC and IVAC are needed to answer this question and results from the present study may help to improve the design of such studies. In a recent post-hoc analysis, VAP clinical practice guideline implementation – with increasing overall guideline compliance - modestly decreased VAC rates but not IVAC. No specific preventive measure alone was associated with lower VAC rates and interventions targeted at VAE specifically were not evaluated[25].

In the present study, the great majority of VACs were identified by increases in PEEP as opposed to FiO_2 increases. In our ICUs we have implemented the higher PEEP/lower FiO_2 protocol from the ARDSnet guidelines[28]. Using a decreased FiO_2 cutoff improved sensitivity for the detection of VAP with a similar positive predictive value, which, in our setting, may therefore be a preferable alternative to the original algorithm. Importantly, the VAE algorithm was largely unknown during the study period and has not been adopted in the Netherlands, and therefore clinicians' decisions to change ventilator settings are expected to be fully independent of the new algorithm and in compliance with local protocols.

Introduction of the VAE algorithm was driven by a desire for more objective, efficient, and reliable measures of complication of mechanical ventilation[17] than the current manual assessment of VAP occurrence[11]. Although the VAE algorithm uses objective criteria and is amenable to automated implementation, our sensitivity analyses demonstrate that

small modifications in electronic implementation lead to important differences in events detected and estimates of associated mortality. In addition, previous studies have shown that manually collected variables are often different from those collected electronically[29], thus care must be taken when comparing rates collected through manual surveillance with electronic surveillance. Because standardized electronic surveillance is not yet universally implemented, comparability across institutions using different electronic implementations or manual systems remains questionable and these concerns will need to be addressed in the future. Finally, from the perspective of benchmarking, additional developments with regard to case-mix correction are necessary before valid interhospital comparisons can be made.

Limitations

This study has several limitations. The reference standard used in this study, prospective VAP surveillance, is inherently vulnerable to the disadvantages of VAP surveillance described previously. However, the assessment was done prospectively by multiple well-trained assessors and completely independent of the VAE algorithm. A prior study found the agreement between raters was high overall (89%), but lower for VAP (35%)[20]. However, as opposed to the retrospective study setting, the current process of prospective surveillance involves discussions among observers, discussions with (senior) clinicians in multidisciplinary meetings attended by critical care physicians and infection specialists, and continuous checks of data-integrity. All prospective diagnoses were therefore made after consensus. In addition, the reliability of the retrospective study was similar to previously reported studies[11]. Furthermore, although the participating centers did not routinely perform bronchoalveolar lavage in patients with a clinical suspicion of VAP, this is very representative for the diagnostic practices in most ICUs worldwide and thus adds to the generalizability of our findings. In our setting of selective decontamination of the digestive tract (SDD) all patients are screened for respiratory pathogens by collecting endotracheal specimens according to a standardized protocol on admission and at least twice weekly. Respiratory specimens were obtained from all patients suspected of VAP prior to the introduction of antimicrobial therapy and microbiological confirmation was present in all definite or probable cases and in two thirds of possible cases. Second, the SDD regimen potentially lowers the risk of VAP[30,31]. However, we would not expect much better concordance and reliability in settings with higher VAP rates. Finally, the competing-events analysis adjusted for age, gender, admission type, hospital and baseline APACHE as confounders, but did not include time-varying confounders and thus some residual confounding may remain. However, the adjustment methods used were identical for all the entities compared.

CONCLUSIONS

This study shows (i) that detection of VAP by the novel surveillance paradigm for ventilator-associated events is poor, (ii) that events detected as VAE represent a broad range of clinical conditions that may not be liable to preventive measures and thus do not necessarily represent quality of care, and (iii) that small differences in electronic implementation can considerably affect incidence rates and associated mortality of events detected. More studies are needed to establish the clinical entities underlying VAE events, develop methods for case-mix adjustment and ensure that rates obtained from different institutions are comparable prior to considering these events as an established quality metric. Finally, given these important concerns, the VAE paradigm should not be used as a sole method of surveillance for VAP.

Acknowledgements

We would like to acknowledge Cristian Spitoni for statistical advice and Mark de Jong for retrieving mechanical ventilator data.

Members of the MARS Consortium

Friso M. de Beer, M.D., and Lieuwe D. J. Bos, M.D. (Department of Intensive Care Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands); Jos F. Frencken, M.D. (Department of Intensive Care Medicine and Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands); and Roosmarijn T. M. van Hooijdonk, M.D., Mischa A. Huson, M.D., Laura R. A. Schouten, M.D., Marleen Straat, M.D., Lonneke A. van Vught, M.D., Maryse A. Wiewel, M.D., Esther Witteveen, M.D., Gerie J. Glas, M.D., and Luuk Wieske, M.D. (Department of Intensive Care Medicine, Academic Medical Center, University of Amsterdam).

Sources of support

This work was supported by the Center for Translational Molecular Medicine (http://www.ctmm.nl), project MARS (grant 04I-201). MB has received research funding from the Netherlands Organization of Scientific Research (NWO Vici 918.76.611). The sponsors did not play a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

SUPPLEMENTARY DATA

Supplementary methods

Brief descriptions of the VAE algorithm entities and implementation

Ventilator-associated condition (VAC)

Definition: A sustained increase in ventilator requirements after a baseline period of stability or improvement. Either an increase in daily minimum positive end-expiratory pressure (PEEP) of >3 cm H2O or a > 20% increase of the daily minimum fraction of inspired oxygen (FiO2), both sustained for at least two calendar days, are classified as a VAC event. Of note, PEEP settings < 5 are considered equivalent to 5.

Implementation details: Minute-to-minute ventilator settings were extracted from ICU databases. For each day of ventilation (>60 minutes of ventilation), the minimal PEEP and FiO2 settings were used for the VAC algorithm.

Infection-related VAC (IVAC)

Definition: The presence of VAC with the occurrence of hypothermia or fever, leukocytosis or leukopenia and the initiation of new antibiotic therapy for at least four calendar days in the two-day window surrounding the VAC event. The first two days of mechanical ventilation are excluded to prevent inclusion of conditions present on admission.

Implementation: Information on antibiotic use, clinical chemistry results, and clinical signs were extracted from electronic records. Antibiotics considered in the IVAC algorithm were as described by National Healthcare Safety Network (NHSN) with the exception that antibiotics prescribed solely for prophylactic or prokinetic purposes were excluded (routine selective digestive decontamination regimen or low-dose erythromycin). The algorithm was implemented independently by two researchers to ensure correct implementation.

Possible or probable ventilator-associated pneumonia (VAP)

Definition: Microbiological evidence for infection was collected from respiratory tract cultures in the time window surrounding an IVAC. A possible VAP requires purulent respiratory secretions as determined by > 25 neutrophils and < 100 epithelial cells per low power field or a qualitative or semi-quantitative positive culture from a respiratory sample. A probable VAP is diagnosed by purulent respiratory secretions and (semi-)quantitative results from respiratory samples or positive cultures from pleural fluid, histopathology samples or positive diagnostic tests for Legionella spp or respiratory viruses.

Implementation: Assessment of VAP status was performed using microbiology results and sputum characteristics extracted from electronic records.

Source: NHSN. The national healthcare safety network device-associated module: Ventilator - associated event protocol. Atlanta, USA: Centers for Disease Control and Prevention; July 2013.

Supplementary tables

Table S1 | Definition of definite, probable or possible VAP by prospective surveillance (reference standard).

	Possible	Probable	Definite
Clinical criteria*	CPIS > 6	CPIS > 6	CPIS > 6
Radiology	Dubious abnormalities on radiographic examination	Evident abnormalities on radiographic examination†	Evident abnormalities on radiographic examination† OR Radiographic evidence of lung abscess or empyema
Microbiology	Detection of pathogen in semi-quantitative culture from respiratory secretions (endotracheal aspirate or bronchoscopic aspirate)	Detection of pathogen in quantitative culture in BAL (10^4) or PSB (10^3) OR Positive blood culture with pathogen also isolated from airway culture	Histopathologic evidence of pneumonia (abscess with PMN concentration and positive tissue culture) OR If empyema, positive culture of aspirate.

Notes: All events must occur during ICU admission and > 48 hours after onset of mechanical ventilation.
* CPIS – clinical pulmonary infection score (based on tracheal secretions, temperature, leukocyte count, PaO_2/FiO_2 ratio, infiltrates on X-thorax and microbiology). † Evident radiographic abnormalities are defined as new or progressive infiltrates, consolidation, cavitation or pleural effusion.
Abbreviations: BAL – bronchoalveolar lavage, PMN – polymorphonuclear leukocyte, PSB – protected specimen brush. For details: *Klein Klouwenberg PMC, Ong DSY, Bos LDJ, et al. Interobserver agreement of CDC criteria for classifying infections in critically ill patients. Crit Care Med 2013; 41:2373-2378.).*

Table S2 | Incidence of VAE as implemented using the original algorithm by hospital.

	Hospital A	Hospital B
	N (rate/1000 MV days)	N (rate/1000 MV days)
N ICU admissions	1195	1101
N ventilator days	8499	7363
VAC events*	87 (10.2)	71 (9.6)
IVAC events†	33 (3.9)	33 (4.5)
VAE-VAP§	22 (2.6)	29 (3.9)
Prospective VAP‖	60 (7.1)	67 (9.1)

* 158 VAC events in152 patients, † 66 IVAC events in 65 patients, § Includes both possible and probable VAP. 51 events in 50 patients, ‖ Includes possible, probable and definite VAP. 127 events in 115 patients.
Abbreviations: IVAC – infection related ventilator-associated condition, VAC – ventilator-associated condition, VAE – ventilator-associated event, VAP – ventilator-associated pneumonia.

Table S3 | Cause-specific hazard ratios, for each type of event.

	CSHR Death (95% CI)	CSHR Discharge (95% CI)
VAC		
VAC	3.96 (2.43-6.45)*	0.38 (0.26-0.56)*
Age	1.03 (1.02-1.04)*	1.01 (1.00-1.01)
Gender	1.01 (0.83-1.23)	0.99 (0.89-1.09)
APACHE	1.04 (1.04-1.04)*	1.02 (1.02-1.02)*
Hospital	1.15 (0.95-1.38)	0.81 (0.74-0.90)
Medical admission	1.05 (0.85-1.29)	0.96 (0.87-1.06)
VAC*time	0.96 (0.93-0.98)	1.02 (1.01-1.04)
Age*time	1.00 (1.00-1.00)	--
APACHE*time	1.00 (1.00-1.00)	1.00 (1.00-1.00)
IVAC		
IVAC	0.98 (0.57-1.70)	0.47 (0.33-0.66)
Age	1.03 (1.02-1.04)*	1.00 (1.00-1.01)
Gender	1.03 (0.85-1.25)	0.98 (0.88-1.08)
APACHE	1.04 (1.04-1.04)*	1.02 (1.01-1.02)*
Hospital	1.14 (0.95-1.38)	0.83 (0.75-0.91)
Medical admission	1.04 (0.85-1.28)	0.98 (0.89-1.08)
Age*time	1.00 (1.00-1.00)	--
APACHE*time	1.00 (1.00-1.00)	1.00 (1.00-1.00)
VAE-VAP		
VAE-VAP	1.11 (0.60-2.05)	0.56 (0.30-1.05)*
Age	1.03 (1.02-1.04)*	1.01 (1.00-1.01)
Gender	1.02 (0.84-1.24)	0.98 (0.89-1.09)
APACHE	1.04 (1.03-1.04)*	1.02 (1.01-1.02)*
Hospital	1.16 (0.96-1.40)	0.83 (0.75-0.92)
Medical admission	1.82 (1.36-2.44)*	0.98 (0.89-1.09)
VAE-VAP*time	--	0.98 (0.96-1.00)
Age*time	1.00 (1.00-1.00)	--
APACHE*time	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Medical admission*time	0.94 (0.92-0.96)	--
PROSP-VAP		
PROSP-VAP	2.00 (1.34-3.00)	0.45 (0.34-0.58)
Age	1.03 (1.02-1.04)*	1.01 (1.00-1.01)
Gender	1.01 (0.82-1.21)	1.00 (0.90-1.10)
APACHE	1.04 (1.03-1.04)*	1.02 (1.01-1.02)*
Hospital	1.16 (0.96-1.40)	0.83 (0.75-0.91)
Medical admission	1.98 (1.48-2.65)*	0.95 (0.86-1.05)
Age*time	1.00 (1.00-1.00)	--
APACHE*time	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Medical admission*time	0.93 (0.91-0.95)	--

Notes – * Time*exposure interaction included. Abbreviations: APACHE IV – acute physiology and chronic health evaluation IV, CI – confidence interval, CSHR –cause-specific hazard ratio, ventilator-associated SHR – subdistribution hazard ratio, IVAC – infection related ventilator-associated condition, VAC – ventilator-associated condition, VAE-VAP – ventilator-associated pneumonia by the VAE paradigm, PROSP-VAP – ventilator-associated pneumonia by prospective surveillance.

REFERENCES

1. Panzer RJ, Gitomer RS, Greene WH, Webster PR, Landry KR, Riccobono CA. Increasing demands for quality measurement. *JAMA.* **2013**;310:1971-1980.
2. HAI Reporting Laws and Regulations. **2011**. APIC Last accessed: Feb 13 2012.
3. Klevens RM, Edwards JR, Richards CL, Jr. et al. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public Health Rep.* **2007**;122:160-166.
4. Edwards JR, Peterson KD, Mu Y et al. National Healthcare Safety Network (NHSN) report: data summary for 2006 through 2008, issued December 2009. *Am J Infect Control.* **2009**;37:783-805.
5. van der Kooi TI, Mannien J, Wille JC, van Benthem BH. Prevalence of nosocomial infections in The Netherlands, 2007-2008: results of the first four national studies. *J Hosp Infect.* **2010**;75:168-172.
6. Melsen WG, Rovers MM, Groenwold RH et al. Attributable mortality of ventilator-associated pneumonia: a meta-analysis of individual patient data from randomised prevention studies. *Lancet Infect Dis.* **2013**;13:665-671.
7. Skrupky LP, McConnell K, Dallas J, Kollef MH. A comparison of ventilator-associated pneumonia rates as identified according to the National Healthcare Safety Network and American College of Chest Physicians criteria. *Crit Care Med.* **2012**;40:281-284.
8. Uckay I, Ahmed QA, Sax H, Pittet D. Ventilator-associated pneumonia as a quality indicator for patient safety? *Clin Infect Dis.* **2008**;46:557-563.
9. Michetti CP, Fakhry SM, Ferguson PL, Cook A, Moore FO, Gross R. Ventilator-associated pneumonia rates at major trauma centers compared with a national benchmark: a multi-institutional study of the AAST. *J Trauma Acute Care Surg.* **2012**;72:1165-1173.
10. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control.* **2008**;36:309-332.
11. Klompas M. Interobserver variability in ventilator-associated pneumonia surveillance. *Am J Infect Control.* **2010**;38:237-239.
12. Fabregas N, Ewig S, Torres A et al. Clinical diagnosis of ventilator associated pneumonia revisited: comparative validation using immediate post-mortem lung biopsies. *Thorax.* **1999**;54:867-873.
13. Klompas M. Eight initiatives that misleadingly lower ventilator-associated pneumonia rates. *Am J Infect Control.* **2012**;40:408-410.
14. Lin MY, Bonten MJ. The Dilemma of Assessment Bias in Infection Control Research. *Clin Infect Dis.* **2012**;54:1342-1347.
15. Klompas M, Khan Y, Kleinman K et al. Multicenter Evaluation of a Novel Surveillance Paradigm for Complications of Mechanical Ventilation. *PLoS One.* **2011**;6:e18062.
16. National Healthcare Safety Network (NHSN). The National Healthcare Safety Network device-associated module: ventilator-associated event protocol. 2013 Atlanta GA
17. Magill SS, Klompas M, Balk R et al. Developing a new, national approach to surveillance for ventilator-associated events*. *Crit Care Med.* **2013**;41:2467-2475.
18. Klein Klouwenberg, P. M., van Mourik, M. S., Ong, D. S., Cremer, O. L., and Bonten, M. J. Validation of a new surveillance paradigm for ventilator-associated pneumonia [abstract]. **2013**. European Conference of Clinical Microbiology and Infectious Diseases (ECCMID) Berlin.
19. van Mourik MS, Klein Klouwenberg PM, Ong DS et al. Validation and assessment of the new surveillance paradigm for ventilator-associated events [abstract]. *Antimicrob Resist Infect Control.* **2013**;2:O164.
20. Klein Klouwenberg PM, Ong DS, Bos LD et al. Interobserver agreement of Centers for Disease Control and Prevention criteria for classifying infections in critically ill patients. *Crit Care Med.* **2013**;41:2373-2378.
21. Fine JP, Gray RJ. A proportional hazard model for the subdistribution of competing risk. *J Am Stat Assoc.* **1999**;94:496-509.
22. Beyersmann J, Gastmeier P, Wolkewitz M, Schumacher M. An easy mathematical proof showed that time-dependent bias inevitably leads to biased effect estimation. *J Clin Epidemiol.* **2008**;61:1216-1221.

23. Beyersmann J, Allignol A, Schumacher M. *Competing Risk and Multistate Models with R*. New York: Springer; **2012**.
24. Hayashi Y, Morisawa K, Klompas M et al. Toward Improved Surveillance: The Impact of Ventilator-Associated Complications on Length of Stay and Antibiotic Use in Patients in Intensive Care Units. *Clin Infect Dis.* **2013**;56:471-477.
25. Muscedere J, Sinuff T, Heyland DK et al. The Clinical Impact and Preventability of Ventilator-Associated Conditions in Critically Ill Mechanically Ventilated Patients. *Chest.* **2013**.
26. Ding S, Kilickaya O, Senkal S, Gajic O, Hubmayr RD, Li G. Temporal Trends of Ventilator-Associated Pneumonia Incidence and the Effect of Implementing Health-care Bundles in a Suburban Community. *Chest.* **2013**;144:1461-1468.
27. Klompas M, Magill S, Robicsek A et al. Objective surveillance definitions for ventilator-associated pneumonia*. *Crit Care Med.* **2012**;40:3154-3161.
28. ARDSNet. Mechanical ventilation protocol summary of the NIH NHLBI ARDS clinical network. **2014**. http://www.ardsnet.org/system/files/Ventilator%20Protocol%20Card.pdf Last accessed: Mar 1 2014.
29. Bosman RJ, Oudemane van Straaten HM, Zandstra DF. The use of intensive care information systems alters outcome prediction. *Intensive Care Med.* **1998**;24:953-958.
30. Liberati A, D'Amico R, Pifferi S, Torri V, Brazzi L, Parmelli E. Antibiotic prophylaxis to reduce respiratory tract infections and mortality in adults receiving intensive care. *Cochrane Database Syst Rev.* **2009**;CD000022.
31. Bonten MJ. Ventilator-associated pneumonia and the gastropulmonary route of infection: a pendulum. *Am J Respir Crit Care Med.* **2011**;184:991-993.

7

Case-mix adjustment for comparison of outcomes in mechanically ventilated patients using electronic routine care data

Maaike S.M. van Mourik[a,b], Karel G.M. Moons[c], Michael V. Murphy[b], Marc J.M. Bonten[a,c§], Michael Klompas[b,d§] – for the CDC Prevention Epicenters
§ both authors contributed equally.

[a] *Department of Medical Microbiology, University Medical Center Utrecht*
[b] *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute, Boston*
[c] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht*
[d] *Department of Medicine, Brigham and Women's Hospital, Boston*

Manuscript under revision

ABSTRACT

Objective – Healthcare-associated infection rates are increasingly used for benchmarking of hospitals and pay-for-performance incentives. Valid comparisons between healthcare institutions, however, require accurate and efficient correction for disease severity and other differences in hospitals' case-mix. Existing risk adjustment models tend either to be very simplistic or to require extensive manual data collection. We aimed to develop a disease severity prediction model based solely on data routinely available in electronic health records for the purpose of group-level risk adjustment in mechanically ventilated patients.

Design – Retrospective cohort study.

Patients and setting – 20,028 mechanically ventilated patients in a single tertiary medical center (2006 – 2012), of whom 3027 deceased in hospital.

Measurements and Main Results – Predictors were extracted from electronic data repositories (demographics, laboratory tests, medications, microbiology results, procedure codes and comorbidities) and evaluated for suitability for inclusion in a prediction model. Models of increasing complexity for in-hospital mortality were built using logistic regression; internal validation was performed by bootstrapping (bias-corrected area under the ROC curve 0.83-0.88). A simple model including demographics, type of ICU, time to intubation, blood culture sampling, eight common laboratory tests and surgical status achieved an area under the ROC curve of 0.87 (95% CI 0.86-0.88) with adequate calibration. The predicted risk of mortality was associated with occurrence of events detected by the recently introduced ventilator-associated events surveillance paradigm.

Conclusions – Accurate prediction of mortality risk in ventilated patients using electronic, routine care data was feasible using simple models. These estimates of disease severity may be useful for adjustment of ventilator-associated event rates. Future research is necessary to validate and refine these models and to further assess their suitability to risk-adjust ventilator-associated events.

INTRODUCTION

Evaluating quality of care has become increasingly important over the past decades. Mandatory public reporting and non-payment programs for medical complications have strengthened the need for reliable and comparable measures of performance[1,2]. Mandatory reporting of infection rates has been introduced in a majority of the United States as well as other countries, and regulations for non-payment for infections considered preventable are expanding[3-5]. Importantly, this emphasis on transparency of quality of care measured by patient outcomes requires not only reliable ascertainment of clinical outcomes but also accurate and practical methods of case-mix adjustment for underlying risk as to allow for valid comparisons between healthcare facilities[1,6-10].

Recently, a new surveillance paradigm for complications of mechanical ventilation was introduced as an alternative to traditional assessment of ventilator-associated pneumonia. This novel approach aims to simplify surveillance, decrease subjectivity, and facilitate electronic implementation[11]. Conditions identified by this new algorithm are reportable to the Centers for Disease Control's National Healthcare Safety Network (NHSN). Two tiers of the new framework - ventilator-associated conditions (VAC) and infection-related VAC (IVAC) - are being considered for future use in benchmarking and pay-for-performance programs[11]. Currently, case-mix adjustment is limited to stratification by type of intensive care unit (ICU)[12,13]. Better methods are needed to facilitate more meaningful interpretation and comparison of rates.

Several intensive care unit risk scoring systems, such as the *Acute Physiology and Chronic Health Evaluation* (APACHE) and *Simplified Acute Physiology Score* (SAPS) have been in use for numerous years to perform case-mix adjustment in research and outcome evaluation[8,14-16]. Although these models have good predictive accuracy for in-hospital mortality, their universal adoption is hampered by the requirement for extensive collection of clinical parameters such as admission diagnoses, Glasgow Coma Scale scores, urine output, mean arterial blood pressure and other physiological parameters. These variables are labor intensive to collect and infrequently available in a structured electronic format[17,18]. With the increasing use of electronic health records (EHRs), there is growing interest in the possibility of using electronic data as a reliable alternative to time-consuming manual data collection[19-21]. Some risk scoring systems have been adapted to electronic data[19], however their generalizability is questionable given their use of diagnosis codes from administrative data, inclusion of clinical variables that are only available in a fraction of comprehensive EHRs, and the incorporation of some physiological variables and laboratory tests that are not routinely measured in all patients[19-21]. Furthermore, existing risk-adjustment models may not be optimal for ventilator-associated event

(VAE) risk adjustment since they were derived on a mix of ventilated and non-ventilated patients and, in some cases, excluded cardiovascular surgery patients[14-16,20].

In this study, we aimed to develop a disease severity prediction model targeted specifically at the use of electronically stored, routinely collected data for the purpose of group-level risk adjustment in mechanically ventilated patients. We explore how the balance between model precision, complexity and anticipated generalizability of data collection can be optimized to achieve less complex yet precise estimates of severity of disease. We also assessed the relation between estimated mortality risk using our prediction model and the risk of VAEs.

MATERIALS AND METHODS

Study population

All consecutive patients initiated on mechanical ventilation (MV), either through an endotracheal tube or tracheostomy, and admitted to an ICU on mechanical ventilation in the Brigham and Women's Hospital (Boston, MA) between January 1st 2006 and December 31st 2012 were eligible for study inclusion. An interruption of MV of more than 1 complete calendar day defined the start of a new ventilation episode. Only the first episode within a hospital admission was selected, whenever applicable. If multiple hospital admissions occurred for a patient within the study period, a random admission was chosen. Patients under the age of 18, patients who could not be successfully linked to electronic data repositories and patients who were not admitted to an ICU were excluded. This study was reviewed and approved by the Institutional Review Board of Brigham and Women's Hospital (protocol number: 2012P002052).

Outcome

The primary outcome of this study was in-hospital mortality as estimates of disease severity based on mortality may be used for risk-adjustment for all tiers of the VAE algorithm; in addition, the ongoing periodic revisions to the VAE definitions would require frequent model adaptation. As a sensitivity analysis, models were also developed to predict mortality within 30 days of initiation of MV. Day 30 mortality was determined from hospital vital status records[22-24].

Predictors

Predictors were collected from the hospital's electronic Research Patient Data Registry (RPDR)[25]. This database compiles a broad variety of structured electronic data elements recorded during routine clinical care. We extracted data on patient demographics, laboratory tests, microbiology results, procedure codes, diagnosis codes, billing information and medication use. Variables were selected for extraction based on clinical associations with severity of disease and on previously published risk-adjustment models[14,15,20,26]. Dates on ICU admission and discharge could not be extracted from the RPDR. Subsequently, we assessed all candidate variables for potential inclusion in a mortality prediction model. We made a qualitative estimate of feasibility based on the extent of missing data observed in our sample and the expected likelihood that similar data would be available in electronic form in other hospitals. We also evaluated candidate variables for generalizability based on the probability that predictors would have similar interpretation when applied in other settings or across time (**table 1**). When applicable, possible predictors were classified in broad groups that are expected to translate to other settings. As such, ICU types were grouped into broad categories (medical, surgical, neuroscience and cardiovascular surgery), and procedure types were classified by anatomic/functional location (e.g. muscoskeletal, cranial, digestive tract). We sought to develop a model with high predictive performance using the minimum number of high-quality predictors (i.e. high feasibility and generalizability).

We collected values for predictors from the (calendar) day of MV onset and the day prior when available (not necessarily days of ICU admission). Data were cleaned and processed as specified in section 1A – 1C of the Supplementary material; for processing of laboratory variables this includes an automated algorithm to detect and replace extreme outlier values. For each episode of MV, the observation reflecting the worst prognosis measured on the day of MV onset or the day prior was taken (e.g. lowest hematocrit, highest creatinine) for each predictor. Comorbid conditions were assessed using the Elixhauser method; this system is designed to detect pre-existing comorbidities in administrative data and was implemented using the specifications provided by the Healthcare Cost and Utilization Project[27,28].

Model development & statistical methods

Descriptive statistics of the study population were generated and data integrity was assessed. Patterns of missing data were explored and multiple imputation was performed using the *mice* package in R (10 imputation sets)[29]. Exploratory univariable analysis of the association between predictors and in-hospital mortality was performed and restricted

Table 1 | Overview of electronic data sources - collected during the process of routine care on the first calendar day of mechanical ventilation or the day prior when available. All groups of predictors are classified by anticipated feasibility and generalizability of electronic data collection.

Predictor category	Description	Expected feasibility[b]	Expected generalizability[c]	% Missing (Comments)
Demographics	Age, gender	+	+	None
Unit type[a]	Medical, surgical, neuroscience or cardiovascular surgery	+	+	None
Time to ventilator	Days from hospital admission to start of mechanical ventilation	+	+	None
Microbiology	Any blood culture taken (regardless of result)	+	+	None
Procedure codes	Classified in broad categories: Muscoskeletal, thorax, CABG, valve, heart/lung transplant, aneurysm, digestive tract, spleen, cranial.	+	+	None
Physiology I	Hematocrit, white blood cell count, platelets, creatinine, glucose, sodium, international normalized ratio (INR)	+	+	1.2%
Physiology II	Albumin, total bilirubin, ALT	+/−	+/−	40.0%
Physiology III	Blood gas analysis (pH, pO2, pCO2)	+/−	+/−	6.9%
Medication	(Groups of) medications initiated: Antibiotic (2 or more), antifungal, amiodarone, number of vasopressors, anticonvulsant, steroid, vitamin K, airway medication, antipsychotic.	+/−	−	None (Dependent on clinical practice)
Comorbidities	Elixhauser comorbidity groups (based on ICD9 and DRG codes)	+/−	−	None (Dependent on coding practices)

[a] Surgical ICU: all surgical ICU types (general, trauma, thoracic, …) Neuroscience ICU: neurology and neurosurgery. [b] Feasibility was estimated based on the amount of missing data in our sample and the expected availability from electronic health records in other settings. [c] Generalizability was defined as the anticipated consistency of predictor interpretation in other settings.
Notes: +: good anticipated feasibility/generalizability, +/−: some concerns, −: limited anticipated feasibility/generalizability. For details on variable definitions please refer to supplementary material (section 1).
Abbreviations: ALT – alanine transaminase, CABG – coronary artery bypass graft, DRG – diagnosis-related group, ICD9 – International classification of disease, 9th edition, INR – international normalized ratio, MV – mechanical ventilation.

cubic splines were used for continuous variables that did not have a linear association with the log-odds of mortality[30] (Supplemental material, section 1B).

Regression models of incremental complexity were built, first including candidate predictors with the best expected feasibility and generalizability. All variables within each group of predictors were added to the model irrespective of their individual significance in univariable or multivariable analysis (e.g., no backward or forward selection of main effects). Interactions were judged relevant if $p < 0.0001$ and were detected by backward elimination. If a group of predictors as a whole (as defined in **table 1**) no longer contributed to model performance as measured by Akaike's Information Criterium it was eliminated in its entirety to reduce model complexity. To limit overfitting and obtain bias-corrected estimates of model performance, internal validation followed by uniform shrinkage was performed using bootstrap samples (100 samples per multiple imputation set) as this method generates more accurate estimates of performance than split-sample or cross-validation approaches[31]. Parameter estimates from the imputation sets were pooled by Rubin's rule[32].

Model discrimination and calibration were assessed using the area under the receiver-operator characteristic (ROC) curve, the Brier score, the Hosmer & Lemeshow test, and examination of calibration plots. Given the large heterogeneity in mortality across ICU types, the model with the most favorable balance of complexity to performance was also employed to develop separate models by ICU type. Finally, the relationship between severity of underlying disease, as measured by predicted risk of mortality obtained from the models developed stratified by ICU type, and observed incidence of VAEs was assessed. For the overall population and for each type of ICU separately, patients were classified by decile of predicted mortality risk and incidence of VAEs was determined within each decile. All analyses were performed using SAS 9.2 (Cary, NC) and R (www.r-project.org).

RESULTS

During the study period, 24,366 episodes of MV were recorded by the department of respiratory therapy. Among the 20,581 episodes in unique patients, 59 involved patients under the age of 18, a successful medical record match was not possible in 350 and 144 occurred in patients that were not admitted to an ICU, thus leaving 20,028 episodes for analysis. Median age of the patients included was 64 years, 60.3% were male and about 40% of the patients were admitted to the cardiovascular surgery unit (**table 2**). Average in-hospital mortality was 15.1%, ranging from 3.3% for cardiovascular surgery to 38% for the medical ICUs.

Table 2 | Selected patient characteristics.

	2006 – 2012	Discharged alive	Deceased in hospital
Median (IQR) or n (%)	**n = 20028**	**n = 17001**	**n = 3027**
Age (years)	64 (53-73)	64 (52-74)	67 (56-78)
Male	12084 (60.3)	10362 (61.0)	1722 (56.9)
Unit type			
Medical	4282 (21.4)	2663 (15.7)	1619 (53.5)
Surgical	5347 (26.7)	4720 (27.8)	627 (20.7)
Neuroscience	2252 (11.2)	1742 (10.3)	510 (16.9)
Cardiovascular surgery	8147 (40.7)	7876 (46.3)	271 (9.0)
Time to ventilation onset			
Day of hospital admission	8507 (42.5)	7333 (43.1)	1174 (37.8)
Day 1 or 2 after admission	7097 (35.4)	6257 (36.8)	840 (27.8)
Days 3-7 after admission	3017 (15.1)	2479 (14.6)	538 (17.8)
≥ 8 days after admission	1407 (7.1)	932 (5.5)	475 (15.7)
Any surgical procedure[a]	13457 (67.2)	12949 (73.5)	963 (31.8)
Blood culture obtained[a]	4315 (21.5)	2860 (16.8)	1455 (48.1)
Duration of mechanical ventilation (days)	2 (1-3)	2 (1-3)	3 (2-8)
Length of hospital stay (days)	11 (7-18)	11 (8-18)	9 (4-19)

[a] On day of mechanical ventilation onset or day prior.
Abbreviations: IQR – interquartile range, MV – mechanical ventilation.

The result of model development - by incremental complexity - is presented in **table 3.** As expected, models using larger numbers of predictors have improved performance characteristics, although the gains with incrementally more complicated models become smaller. Model 2, using only demographic information, unit type, microbiology data, procedure codes and commonly performed laboratory tests achieved an optimism-corrected area under the ROC curve of 0.870 (95% CI 0.864-0.876). Parameter estimates are presented in the supplemental material (section 2, **table S1**). Internal validation showed that over-optimism was very limited; the area under the ROC curve was only 0.003-0.006 lower than the apparent area under the ROC curve. Examination of calibration plots showed adequate calibration with slight over-prediction for high-risk patients (supplementary material, **figure S1**); this is also reflected by the Hosmer-Lemeshow test statistics (X^2 ranging from 49.9 to 60.2 on 10 degrees of freedom for models 2 to 5). A sensitivity analysis for 30-day mortality demonstrated similar performance of all models (supplemental material, **table S2**).

Model 2 also had reasonable performance when applied within ICU types (area under the ROC curve ranging from 0.758 to 0.852); as can be expected the observed performance is somewhat lower as differences between patients within each of the unit types are smaller than when comparing across units. Given the large heterogeneity in mortality rates between types of ICUs and the intended use of achieving adequate risk-adjustment

Table 3 | Results of model development and internal validation for in-hospital mortality. Models are ordered by increasing complexity as classified by feasibility and generalizability of data collection.

Predictor category	Model 1	Model 2	Model 3	Model 4	Model 5
Demographics	x	x	x	x	x
Unit type	x	x	x	x	x
Time to vent	x	x	x	x	x
Microbiology[a]	x	x			
Procedure codes	x	x	x	x	x
Physiology I[b]		x	x	x	x
Physiology II[b] + III[b]			x	x	x
Medication				x	x
Comorbidities					x
N independent predictors	17	36	47	56	85
Area under ROC curve[c]	0.830	0.870	0.877	0.879	0.884
(95% CI)	(0.823-0.837)	(0.864-0.876)	(0.871-0.883)	(0.874-0.885)	(0.878-0.889)
- Medical ICU	0.674	0.758	0.775	0.783	0.797
- Surgical ICU	0.736	0.822	0.835	0.840	0.844
- Neuroscience ICU	0.724	0.784	0.786	0.794	0.794
- CV surgery	0.770	0.852	0.860	0.856	0.871
H & L statistic	12.7	50.3	60.2	49.9	56.0
p-value H&L	0.120	<0.001	<0.001	<0.001	<0.001
Brier score	0.102	0.091	0.089	0.087	0.086

[a] Microbiology (blood culture taken) did not contribute to model performance in models 3 and up and was therefore removed. [b] Physiology I–III, please refer to table 1 and supplementary data for definitions.
[c] bias-corrected.
Abbreviations: CV surgery – cardiovascular surgery, df – degrees of freedom, ICU – Intensive care unit, H&L statistic – Hosmer & Lemeshow statistic, ROC – Receiver Operating Characteristic

within ICU types, the model was refit stratified by ICU type (**table 4**, **table S3**) resulting in similar performance characteristics yet less complex models as interaction terms were no longer necessary.

Table 4 shows that depending on the type of ICU, the overall risk of VAC ranges from 0.9% to 9.0% and from 0.4% to 3.2% for IVAC. The association between predicted mortality (model 2, stratified by ICU type) and the occurrence of VAE within each type of ICU is presented in **figure 1**. The top row of the figure demonstrates a positive association between predicted risk of mortality and incidence of VAE across the overall population. However, risk differentiation within ICU types is also possible as shown by the bottom four rows; for surgical and cardiovascular surgery units there is a positive association between predicted risk of mortality and VAC incidence, whereas for medical and neurological ICUs the picture is less clear. In order to be at risk for a VAE, patients must

Table 4 | Model development for in-hospital mortality stratified by unit type for model 2.

	Medical	Surgical	Neuroscience	Cardiovascular surgery
N episodes of MV	4282	5347	2252	8147
N deceased (%)	1619 (37.8)	627 (11.7)	510 (22.6)	271 (3.3)
N VAC (/100 MV episodes)	390 (9.1)	402 (7.5)	112 (5.0)	67 (0.9)
N IVAC (/100 MV episodes)	135 (3.2)	167 (3.1)	51 (2.3)	33 (0.4)
VAC rate/1000 MV days	16.1	15.6	10.4	4.3
IVAC rate/1000 MV days	5.6	6.5	4.8	2.1
Area under ROC curve[a]	0.754	0.818	0.772	0.856
(95% CI)	(0.740-0.769)	(0.802-0.835)	(0.750-0.793)	(0.835-0.876)
H & L statistic	10.6	21.9	8.3	12.8
Brier score	0.189	0.084	0.142	0.027

[a] after correction for over-optimism.

Note: Models were developed by unit type; procedure types that did not occur within the unit types were excluded. There were no significant interactions in all models. Abbreviations: df – degrees of freedom, H & L statistic – Hosmer and Lemeshow statistic, IVAC – Infection-related ventilator-associated condition, MV – mechanical ventilation, ROC – Receiver Operating Characteristic, VAC – Ventilator-associated condition.

by definition have received at least four days of MV – two days of stability or improvement followed by two days of deterioration[33]. The bar charts in the right column of the figure depicts the fraction of patients in each decile of predicted mortality risk that received four or more days of MV and therefore are eligible for a possible VAE. As can be seen, the association between observed risk of VAE and predicted mortality is largely explained by the proportion of patients that remain on MV four days or longer and are thus at risk. In particular for medical units, the decline in VAC incidence in higher mortality risk groups is caused by an increasing proportion of patients dying within four days of MV onset.

DISCUSSION

Assessing quality of care is of increasing importance to stakeholders in healthcare delivery systems and adjustment for severity of underlying disease is a necessary requirement to achieve valid comparisons of patient outcomes across healthcare providers. This study is meant as a proof-of-concept that predicting risk of mortality in mechanically ventilated patients for the purpose of adjusting for differences in severity of underlying disease is feasible using fairly simple models applied to routine care electronic data; the performance characteristics obtained were in the same range as those generated by more labor-intensive methods such as APACHE and SAPS (area under ROC curve 0.82-0.88)[8]. Models of increasing complexity showed incrementally better performance but at the cost of more extensive data collection and rising concerns with generalizability to different settings. Furthermore, there is a clear relation between predicted risk of mortality and the likelihood of VAEs, thus providing a starting point towards improved correction

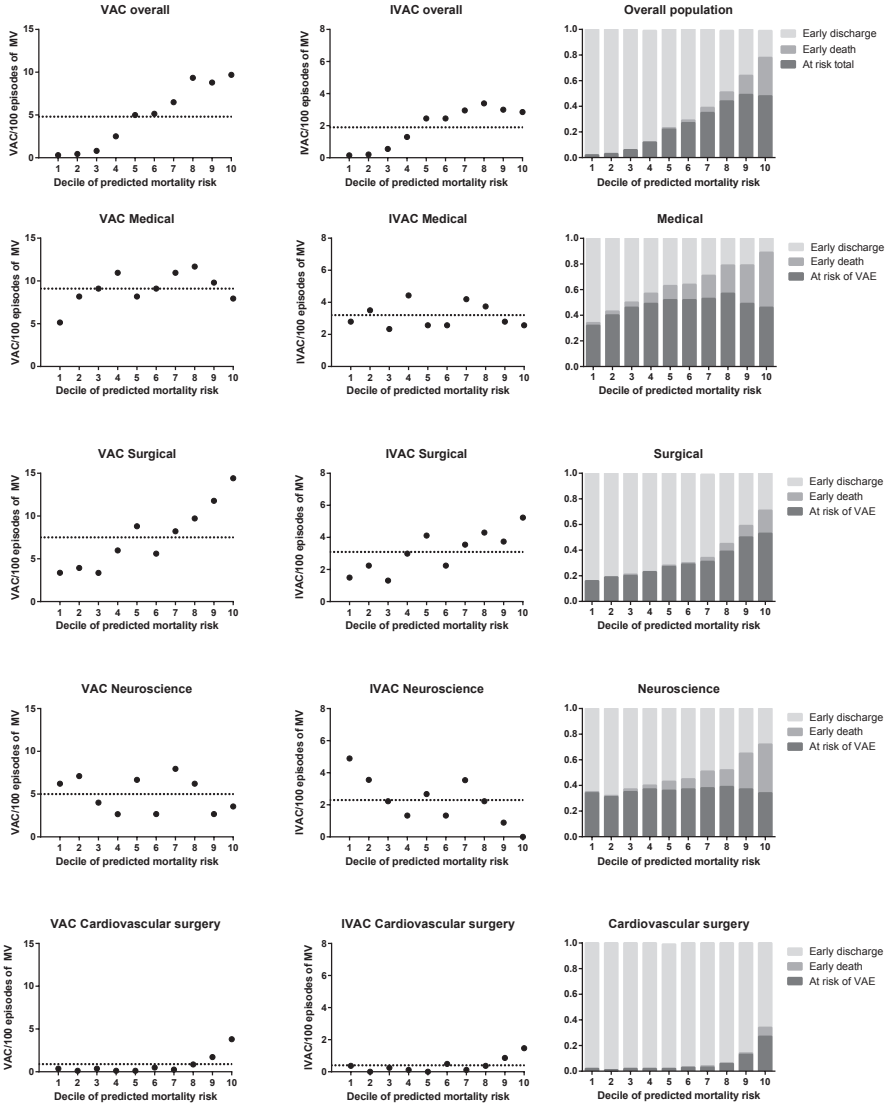


Figure 1 | Association between predicted risk of in-hospital mortality (model 2 stratified by unit type) and incidence of ventilator-associated conditions (VAC, left column) and infection-related VAC (IVAC, middle column).

Horizontal lines represent the overall event incidence. The bar charts (right column) depict the proportion of patients at risk of developing a VAC or IVAC and those either discharged alive off the ventilator within four days of ventilator onset (early discharge) or deceased within four days of ventilator onset (early death).
Top row: deciles of predicted mortality risk calculated for the overall population, bottom four rows: patients are stratified by decile of mortality risk within each ICU type.

for differences in patient populations when these outcomes are used for assessment of quality of care. This association seems to be mediated by identifying the proportion of patients that remain on MV long enough to be eligible for a VAE. The models developed in this study provide more precise risk stratification than current methods used by NHSN (stratification by ICU type alone). As the models were developed in a single hospital, future validation studies will need to determine their generalizability to different hospitals and types of ICU. These validation studies will also provide opportunity to refine estimates for a broader range of (specialized) ICU types.

Although numerous models have previously been developed to predict in-hospital mortality, these were not developed for the purpose of patient outcome comparison in the specific domain of mechanically ventilated patients, and several concerns limit their application in risk-adjustment for VAEs including the highly divergent degrees of sophistication in EHR contents across hospitals[14,15,26]. Current models require extensive amounts of manual data collection, including physiologic variables that are not routinely documented in EHRs (e.g. Glasgow coma scale, blood pressure, urine output), and reason for admission, which is subject to considerable inter-rater variability[18,34]. Recently developed electronic risk-adjustment methods make extensive use of administrative data that may not have a comparable interpretation in other settings, within and outside of the United States, and may be vulnerable to transitions to new coding systems such as the 10th revision of the International Classification of Disease. Furthermore, models in some cases include administrative data collected after the first two days of admission, this may better reflect the evolution of disease within hospitals rather than patients' presenting conditions[19,20]. Moreover, there is evidence that data collected manually cannot be compared to electronically collected data when translating a model directly to an electronic environment, as was done by Liu et al.[19,35]. The models developed in this study do have similarities with existing risk-adjustment models, in particular the model developed within the Veteran's Affairs (VA) system by Render et al.[20]. However, this VA model makes extensive use of admission diagnoses and procedures codes and includes more laboratory variables than the simpler models developed in our study. The incremental improvements obtained by adding more data components to models was also observed in other studies developing risk-adjustment systems[36,37].

The process of model development in this study was targeted specifically for use with electronic data and an important component of model development was the a priori assessment of the feasibility and generalizability of predictor categories from the perspective of data collection through EHRs. Preference was given to measuring information 'close to the patient' as opposed to using proxies such as administrative coding data wherever

possible. We found that simple models without admission diagnoses, disease codes or medication use, variables for which use and classification are more likely to differ across settings, were still able to perform well. Importantly, inclusion of ICU type as a predictor requires assumptions to be made regarding similarity of patient characteristics across ICU types in different hospitals. Future validation studies will need to assess robustness to in admission policies and define methods to handle specialty ICUs; perhaps that in further refinements of the model, ICU type classification may be aligned with the current system employed by NHSN. We further incorporated methods for handling of outlier data and approaches to handling of missing observations other than imputing a normal value as is commonly done. Previous research has shown that simply imputing normal values weakens the reliability of risk adjustment[38].

This study found an intricate relation between predicted severity of disease and the incidence of VAEs. Currently, event rates reported to NHSN are stratified by unit type for comparison[13]. Unit type does explain an important fraction of the observed variation in VAE incidence (**table 4**); however, severity of underlying disease measured by predicted risk of mortality was associated with the proportion of patients developing a VAE even within ICU types. The risk of VAEs seems to be most closely mediated by the fraction of patients ventilated for ≥4 days who are therefore at risk for an event. Interestingly, the association between risk of mortality and VAE is most apparent for cardiac surgery and general surgery ICUs compared to medical and neuroscience ICUs. In the latter two, the number of patients at risk for developing VAE is fairly constant, but the fraction not at risk due to early death on the ventilator increases with ascending risk of mortality. This study purposefully included all patients initiated on mechanical ventilation regardless of whether they achieve four days of MV; not only are all days of MV counted in the denominator of VAE rates, the duration of MV is also unknown at the time of ventilation onset (the moment of prediction). Future research is needed to determine how early deaths should be accounted for in VAE risk prediction models, to assess the exact relationship between severity of underlying disease and the likelihood of developing VAE, and to develop practical methods of risk-adjustment.

Several important limitations need to be addressed. Firstly, this study did not include a separate validation sample; use of a random split-sample for validation was judged not to provide sufficient new insight in comparison to the internal validation strategy used[31]. Examination of the bias-corrected performance estimates showed minimal over-optimism. Furthermore, this model was developed in a single hospital that may not be reflective of general U.S. hospitals. Therefore, these models need to be validated and perhaps updated using data from multiple settings to better understand their performance characteristics

and utility, then also accounting for random differences between centers. Nevertheless, the results of this study will inform the design and data collection of such future studies. Initial examination of Hosmer and Lemeshow statistics indicates miscalibration that is statistically significant. Yet, this test is known to be sensitive to large sample sizes and assessment of calibration plots showed adequate calibration[39].

We were not able to perform a head-to-head comparison of this model with other, existing, methods (e.g. APACHE or SAPS) since these severity of disease estimators are not routinely collected in our setting; their absence was the driving force behind the current study. It has been argued that using 30-day mortality is more reliable than in-hospital mortality for the purpose of outcome comparison (and perhaps also for risk-adjustment)[23,24], however data on in-hospital mortality is often collected more consistently than 30-day mortality. In this study, model performance characteristics were similar for both although misclassification for 30-day mortality cannot be excluded. Prior studies have identified several possible predictors that could not be collected from our EHRs but that may be of predictive value in future studies, including (hospital) location prior to ventilator onset, time to ICU admission, transfer status, urgency of admission and/or surgery and the presence of do-not-resuscitate orders[14,15,20,40]. The use of more sophisticated modeling strategies such as neural networks and supervised machine learning have been suggested in the recent literature as efficient methods of model development[21] and focusing these developments on using parameters that can be collected reliably and efficiently would be of added value.

CONCLUSIONS

Valid comparison of patient outcomes and occurrence of VAEs in mechanically ventilated patients requires adequate methods of correction for severity of underlying disease. This study demonstrates the feasibility of predicting mortality in ventilated patients specifically using electronic routine care data and provides insight into the balance between complexity of data collection and model performance. Estimates of severity of underlying disease obtained from these models show some utility in future adjustment of ventilator-associated event rates although further research is necessary to validate and refine the proposed models and to develop more robust methods of benchmarking for ventilator-associated events.

Acknowledgements

We would like to thank Dr. P Zuithoff for his statistical advice.

Conflicts of interest & Sources of funding

KGMM gratefully acknowledges financial contribution by the Netherlands Organization for Scientific Research (project 9120.8004 and 918.10.615), MK has received a grant from the Centers for Disease Control and Prevention (Grant 1U54CK000172). For the remaining authors, none were declared.

SUPPLEMENTAL MATERIAL

Section 1: Variable specifications

A. Procedure codes
B. Laboratory information
C. Medications

Section 2: Model results

Table S1: Parameter estimates for model 2, after internal validation and uniform shrinkage.
Figure S1: Calibration plots of models developed by incremental complexity.
Table S2: Sensitivity analysis for day 30 mortality, by model of increasing complexity
Table S3: Parameter estimates for models developed stratified by ICU type (model 2).

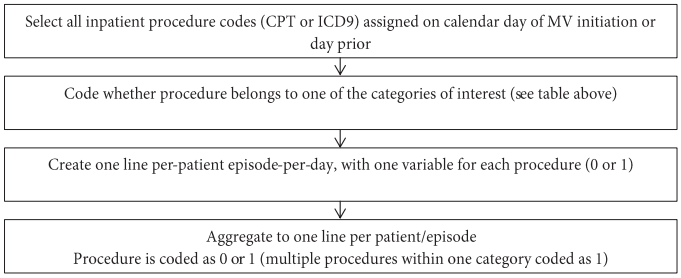
Section 1: Variable specifications

A. Procedure codes

Classification of procedure categories

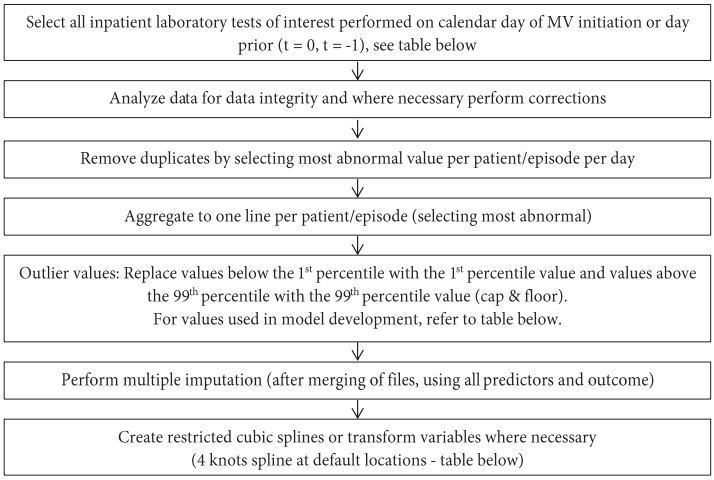
Category	Description, notes	Codes included
Muscoskeletal	Procedure to muscoskeletal system. Excludes casts, strapping	CPT: 20000-28899, 29800-29999, ICD9: 21, 03.53, 76-84
Thorax	Procedure to thorax: lungs or mediastinal space. Excludes tracheostomy, bronchoscopy, lung or heart transplantation, thoracocentesis	CPT: 32000-32850, 32900-32999, 39000-39599. ICD9: 32-33.1, 33.3-33.49, 33.7-34.91, 34.92-34.99.
Heart or Lung transplant	Heart or Lung transplantation	CPT: 32850-32856, 33945, 33935 ICD9: 33.5, 37.51, 33.5
Valve	Valve replacement procedure	CPT: 33400-33499 ICD9: 35.0-35.3
CABG	Coronary artery bypass graft	CPT: 35500-33573 ICD9: 36.1
Aneurysm	Aneurysm procedure (irrespective of location)	CPT: 33860-33909, 34800-35179 ICD9: 39.71, 39.73, 39.79, 38.43-38.49
Procedure of digestive tract	Procedure to the esophagus, stomach, intestine, liver, biliary system, pancreas, anus, abdominal wall. Incl. (diagnostic) endoscopies, gastrostomy and other types of stoma.	CPT: 43020-43659, 43800-50000. ICD9: 17.0-17.3, 42-54 (excl 43.1)
Spleen	Procedure to spleen or splenectomy	CPT: 38100-38201 ICD9: 41.2, 41.4-41.5
Cranial	Procedure to skull or brain, including burr holes	CPT: 61000-62257 ICD9: 1-2, ICD9 7.5-7.7

Flowchart processing of procedure information:



B. Laboratory information

Flowchart processing of laboratory information:



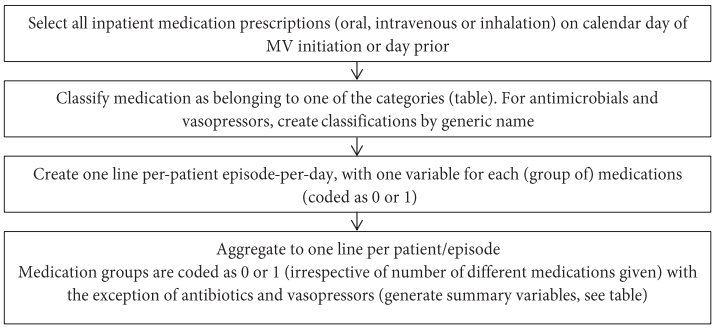
Parameter specifications laboratory variables

Parameter	Unit	Lowest or highest[a]	P1;P99 (cap & floor)	Spline knot locations if applicable[b]
Physiology I				
Hematocrit	%	Lowest	16.3; 45.1	20; 28.2; 40 (3 knots)
White blood cell count	10^3/uL	Lowest	0.61; 29.65	3.15; 7.29; 10.43; 19.13
Platelet count	10^3/uL	Lowest	15; 588	51; 130; 193; 367
Creatinine	mg/dL	Highest	0.39; 7.06	0.57; 0.88; 1.17; 3.19
Blood urea nitrogen	mg/dL	Highest	6; 114	--
Glucose	mg/dL	Lowest	35; 236	63; 94; 111; 164
Sodium	mmol/L	Highest	129; 154	134; 139; 141; 147
INR	-	Highest	1.0; 5.1	--
Physiology II				
Albumin	g/dL	Lowest	1.4; 4.7	--
SGPT	U/L	Highest	5; 1529	--
Total bilirubin	mg/dL	Highest	0.2; 13.6	--
Physiology III				
pO2	mmHg	Lowest	37; 418	55; 92; 127.3; 265;
pCO2	mmHg	Lowest	19; 62	26; 34; 38; 47
pH	--	Lowest	7.0; 7.5	7.15; 7.3; 7.35; 7.44

[a] For all physiology variables, the values most likely associated with a poor patient outcome was selected; depending on the specific variable this can be the lowest or the highest value.
[b] Software to determine knot locations and create spline variables obtained from the rms() package in R, and SAS macros from: http://biostat.mc.vanderbilt.edu/wiki/Main/SasMacros, in particular %rcspline() (last accessed 11-11-2013). Abbreviations: INR – international normalized ratio, SGPT - serum glutamic-pyruvic transaminase (or ALAT), pO2 = partial oxygen pressure, pCO2 = partial carbon dioxide pressure.

C. Medications

Flowchart processing of medication information



Medications included in prediction models.

Medication as predictor	**Coded as:**
Antibiotic	Codes as 2 or more started yes/no (classified by generic active compound)
Antifungal	Yes/no
Amiodarone	Yes/no
Vasopressor	Nr of different pressors started
Anticonvulsant	Yes/no
Steroid	Yes/no
Vitamin K	Yes/no
Airway medication (inhaled)	Yes/no
Antipsychotic	Yes/no

Notes: for all medications, only medication started on t = -1 or t = 0 are included.

Section 2: Model results

Table S1 | Parameter estimates for model 2, after internal validation and uniform shrinkage.

	Estimate	Standard error	Odds ratio	Lower 95% CI	Upper 95% CI
Intercept	9,242		Na	Na	Na
Age	0,026	0,002	1,026	1,023	1,030
Female	-0,040	0,051	0,961	0,870	1,061
Unit group (Medical = reference)					
- Surgical	-0,556	0,112	0,573	0,461	0,714
- Neurology	0,645	0,103	1,906	1,557	2,333
- CV surgery	-1,879	0,221	0,153	0,099	0,235
Time to MV (day of admission = reference)					
- Day 1 or 2 after admission	0,154	0,092	1,166	0,974	1,397
- Days 3-7 after admission	0,747	0,112	2,111	1,696	2,627
- >= 8 days after admission	0,861	0,122	2,366	1,862	3,005
Blood culture (obtained)	0,184	0,057	1,202	1,075	1,345
Procedures (>1 possible):					
Muscoskeletal	-0,724	0,110	0,485	0,391	0,601
Thorax	0,274	0,188	1,315	0,910	1,902
Lung or heart transplant	-1,638	0,373	0,194	0,094	0,404
Cardiac valve	-1,034	0,189	0,356	0,245	0,515
CABG	-1,228	0,211	0,293	0,194	0,443
Aneurysm	-0,582	0,169	0,559	0,401	0,778
Digestive tract	-0,650	0,135	0,522	0,401	0,680
Cranial	0,887	0,451	2,428	1,003	5,874
Laboratory variables					
Hematocrit (min)	0,011	0,011	1,011	0,990	1,032
Spline	-0,007	0,014	0,993	0,967	1,020
White blood cells (min)	-0,079	0,023	0,924	0,884	0,966
Spline 1	0,495	0,110	1,640	1,322	2,035
Spline 2	-1,279	0,307	0,278	0,153	0,508
Platelets (min)	-0,012	0,001	0,988	0,986	0,991
Spline 1	0,032	0,006	1,033	1,020	1,046
Spline 2	-0,070	0,017	0,932	0,901	0,965
Creatinine (max)	-0,826	0,296	0,438	0,245	0,782
Spline 1	22,757	4,648	$7{,}64x10^{9}$	844483,2	$6{,}92x10^{13}$
Spline 2	-51,049	9,977	$0.48x10^{-23}$	$2.06x10^{-31}$	$02.02x10^{-14}$
INR (max)	0,462	0,035	1,587	1,482	1,701
Glucose (min)	-0,015	0,003	0,985	0,980	0,990
Spline 1	0,056	0,012	1,058	1,033	1,082
Spline 2	-0,174	0,043	0,840	0,772	0,915
Sodium (max)	-0,076	0,015	0,927	0,900	0,955

Spline 1	0,130	0,056	1,139	1,021	1,270
Spline 2	-0,157	0,299	0,855	0,475	1,536
Blood urea nitrogen	0,008	0,002	1,008	1,005	1,011
Interactions					
Surgical * Vent onset day 1-2	0,082	0,142	1,085	0,822	1,434
Neurology * Vent onset day 1-2	-0,518	0,159	0,596	0,437	0,813
CV Surgery* Vent onset day 1-2	0,198	0,208	1,219	0,810	1,833
Surgical * Vent onset day 3-7	-0,214	0,175	0,807	0,573	1,137
Neurology * Vent onset day 3-7	-1,368	0,237	0,255	0,160	0,405
CV Surgery* Vent onset day 3-7	0,226	0,226	1,254	0,806	1,950
Surgical * Vent onset day >= 8	-0,181	0,199	0,834	0,565	1,233
Neurology * Vent onset day >= 8	-1,34	0,350	0,262	0,132	0,520
CV Surgery * Vent onset day >= 8	0,304	0,270	1,355	0,798	2,301
Surgical * thorax procedure	-0,754	0,227	0,470	0,301	0,735
Neurology * thorax procedure	0,238	0,525	1,269	0,453	3,553
CV Surgery * thorax procedure	0,632	0,263	1,881	1,124	3,148
Valve procedure* CABG	1,391	0,261	4,019	2,409	6,706
Valve procedure* aneurysm procedure	1,109	0,274	3,031	1,772	5,187
Surgical * Digestive tract procedure	0,180	0,169	1,197	0,860	1,666
Neurology * Digestive tract procedure	-0,813	0,323	0,444	0,236	0,835
CV Surgery * Digestive tract procedure	1,570	0,317	4,807	2,583	8,944
Surgical * Cranial procedure	0,216	0,493	1,241	0,473	3,259
Neurology * Cranial procedure	-1,295	0,465	0,274	0,110	0,682

Abbreviations: CABG – Coronary Artery Bypass Graft, CV Surgery – cardiovascular surgery, MV – mechanical ventilation. Spline knot locations and other data processing details are presented in section 1 of the supplementary data.

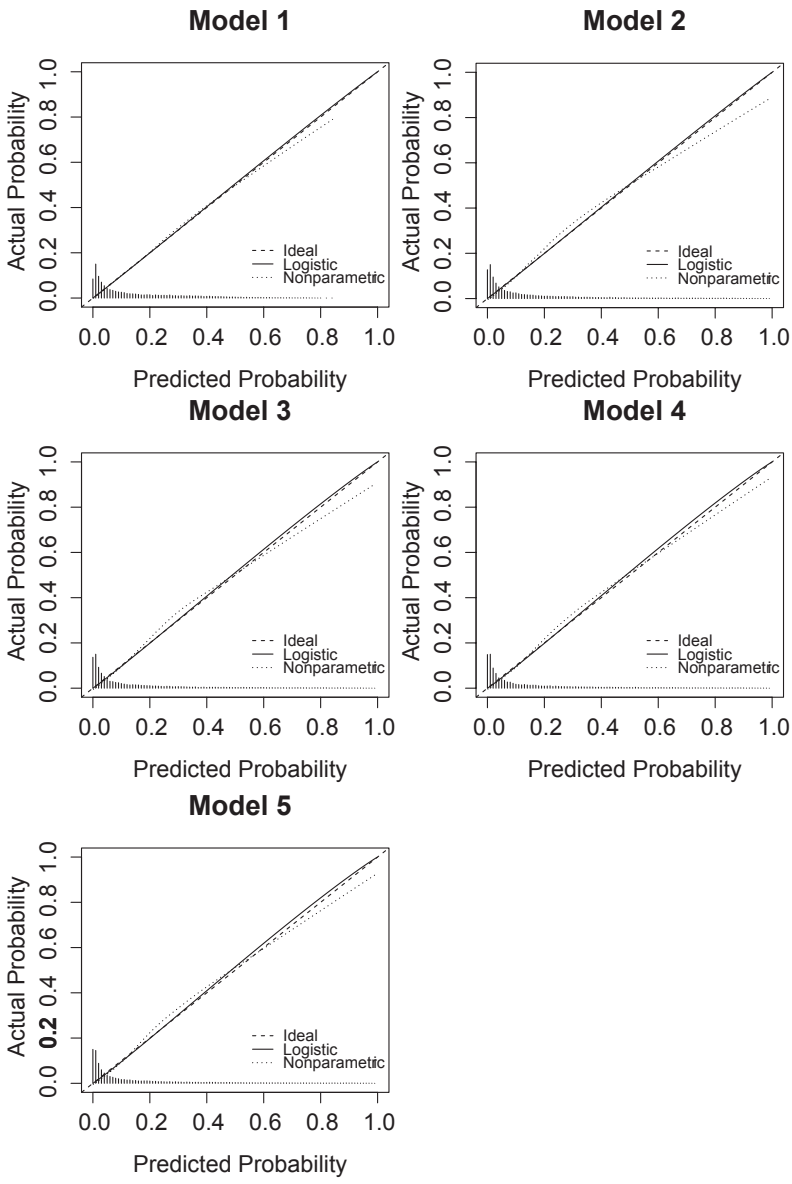


Figure S1 | Calibration plots of models developed by incremental complexity.

Table S2 | Sensitivity analysis for day 30 mortality, by model of increasing complexity.

Predictor category	Model 1	Model 2	Model 3	Model 4	Model 5
Demographic	x	x	x	x	x
Unit type	x	x	x	x	x
Time to vent	x	x	x	x	x
Microbiology[a]	x	x			
Procedure codes	x	x	x	x	x
Physiology I[b]		x	x	x	x
Physiology II[b]+III[b]			x	x	x
Medication				x	x
Comorbidities					x
N independent predictors	17	36	47	56	85
Area under ROC[c]	0.840	0.872	0.877	0.880	0.884
(95% CI)	(0.833-0.847)	(0.866-0.878)	(0.872-0.883)	(0.874-0.886)	(0.879-0.890)
- Medical ICU	0.685	0.752	0.766	0.776	0.792
- Surgical ICU	0.745	0.820	0.826	0.830	0.835
- Neurology ICU	0.736	0.788	0.790	0.795	0.798
- CV surgery	0.762	0.840	0.845	0.846	0.858
H & L statistic	6.6	32.9	51.9	32.8	36.6
p-value H&L statistic	0.580	<0.001	<0.001	<0.001	<0.001
Brier score	0.103	0.094	0.092	0.091	0.089

[a] Microbiology (blood culture taken) did not contribute to model performance in models 3 and up and was therefore removed. [b] Physiology I –III, please refer to table 1 of the main text and supplementary data in section 1 for definitions. [c] bias-corrected.

Abbreviations: CV surgery – cardiovascular surgery, df – degrees of freedom, ICU – Intensive care unit, H&L statistic – Hosmer & Lemeshow statistic, ROC – Receiver Operating Characteristic.

Table S3 | Parameter estimates for models developed stratified by ICU type (model 2).

	Medical		Surgical		Neurology		Cardiovascular surgery	
	Estim.	**SE**	**Estim.**	**SE**	**Estim.**	**SE**	**Estim.**	**SE**
Intercept	7.686	2.749	11,121	4,411	7,462	5,687	-2,873	10,094
Age	0,017	0.002	0,030	0,003	0,039	0,004	0,029	0,006
Female	-0,106	0.730	-0,188	0,105	-0,001	0,126	0,368	0,151
Time to MV (day of admission = reference)								
- Day 1 or 2	0,119	0.091	0,284	0,135	-0,248	0,143	0,065	0,212
- Days 3-7	0,643	0.111	0,580	0,470	-0,447	0,229	0,601	0,225
- >= 8 days	0,830	0.123	0,619	0,179	-0,25	0,371	0,787	0,276
Blood culture (obtained)	0,080	0.076	0,416	0,135	0,043	0,134	0,402	0,287
Procedures (>1 possible):								
Muscoskeletal	-0,940	0.277	-0,549	0,135	-1,348	0,474	0,606	0,479
Thorax	0,194	0.184	-0,348	0,135	0,495	0,509	0,633	0,202
Lung or heart transplant	NA	NA	-1,246	0,47	NA	NA	-1,763	0,686
Cardiac valve	-0,994	0.432	NA	NA	NA	NA	-0,064	0,154
CABG	-0,419	0.560	NA	NA	NA	NA	-0,238	0,149
Aneurysm	-0,669	0.498	-0,548	0,231	-6,948	26,358	0,220	0,208
Digestive tract	-0,650	0.132	-0,436	0,109	-1,361	0,301	0,814	0,310
Cranial	0,752	0.438	1,122	0,212	-0,371	0,126	NA	NA
Laboratory variables								
Hematocrit (min)	0,022	0.016	0,035	0,021	0,016	0,037	0,012	0,032
Spline	-0,041	0.020	0,000	0,028	0,000	0,040	-0,128	0,087
White blood cells (min)	-0,066	0.031	-0,145	0,044	-0,117	0,080	0,021	0,079
Spline 1	0,453	0.156	0,808	0,219	0,616	0,333	0,001	0,387
Spline 2	-1,174	0.437	-2,122	0,614	-1,539	0,892	-0,059	1,118
Platelets (min)	-0,005	0.002	-0,016	0,003	-0,015	0,005	-0,016	0,004
Spline 1	0,003	0.009	0,049	0,014	0,045	0,019	0,059	0,021
Spline 2	0,003	0.025	-0,112	0,036	-0,104	0,050	-0,133	0,062
Creatinine (max)	-1,110	0.434	-0,694	0,597	-0,367	0,676	0,537	1,366
Spline 1	24,390	6.789	18,150	9,387	13,766	11,279	15,316	19,507
Spline 2	-54,06	14.548	-40,54	20,172	-31,09	24,462	-37,62	41,334
INR (max)	0,336	0.045	0,577	0,074	0,458	0,114	0,703	0,110
Glucose (min)	-0,013	0.004	-0,020	0,005	-0,014	0,010	-0,003	0,008
Spline 1	0,042	0.017	0,080	0,024	0,068	0,036	-0,006	0,037
Spline 2	-0,123	0.062	-0,254	0,088	-0,225	0,123	0,137	0,146
Sodium (max)	-0,062	0.020	-0,098	0,032	-0,065	0,041	-0,029	0,073
Spline 1	0,121	0.080	0,225	0,114	0,088	0,139	0,002	0,209
Spline 2	-0,404	0.441	-0,561	0,612	0,104	0,724	0,576	1,024
Blood urea nitrogen	0,008	0.002	0,014	0,004	-0,001	0,006	0,007	0,006
Interactions	None	--	None	--	None	--	None	--

Note: As with all logistic regression models, the odds ratio can be calculated by taking the exponent of the estimated OR = $e^{estimate}$.

Abbreviations: CABG – Coronary Artery Bypass Graft, CV Surgery – cardiovascular surgery, Estim. – Estimate, MV – mechanical ventilation, SE – standard error.

REFERENCE LIST

1. Panzer RJ, Gitomer RS, Greene WH, Webster PR, Landry KR, Riccobono CA. Increasing demands for quality measurement. JAMA. **2013**;310:1971-1980.
2. Rosenthal MB. Nonpayment for performance?Medicare's new reimbursement rule. N Engl J Med. **2007**;357:1573-1575.
3. Centers for Medicare and Medicaid Services. Hospital-Acquired Conditions. **2012**. http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html Last accessed: Nov 25 2013.
4. HAI Reporting Laws and Regulations. **2011**. APIC Last accessed: Feb 13 2012.
5. Haustein T, Gastmeier P, Holmes A et al. Use of benchmarking and public reporting for infection control in four high-income countries. Lancet Infect Dis. **2011**;11:471-481.
6. Mu Y, Edwards JR, Horan TC, Berrios-Torres SI, Fridkin SK. Improving risk-adjusted measures of surgical site infection for the national healthcare safety network. Infect Control Hosp Epidemiol. **2011**;32:970-986.
7. Harris AD, McGregor JC. The importance of case-mix adjustment for infection rates and the need for more research. Infect Control Hosp Epidemiol. **2008**;29:693-694.
8. Breslow MJ, Badawi O. Severity scoring in the critically ill: part 1--interpretation and accuracy of outcome prediction scoring systems. Chest. **2012**;141:245-252.
9. van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. Br J Surg. **2013**;100:628-627.
10. Kritsotakis EI, Dimitriadis I, Roumbelaki M et al. Case-mix adjustment approach to benchmarking prevalence rates of nosocomial infection in hospitals in Cyprus and Greece. Infect Control Hosp Epidemiol. **2008**;29:685-692.
11. Magill SS, Klompas M, Balk R et al. Developing a new, national approach to surveillance for ventilator-associated events*. Crit Care Med. **2013**;41:2467-2475.
12. Centers for Disease Control and Prevention. Surveillance for Ventilator-Associated Events. **2013**. http://www.cdc.gov/nhsn/acute-care-hospital/vae/ Last accessed: Nov 26 2013.
13. Dudeck MA, Weiner LM, len-Bridson K et al. National Healthcare Safety Network (NHSN) report, data summary for 2012, Device-associated module. Am J Infect Control. **2013**;41:1148-1166.
14. Moreno RP, Metnitz PG, Almeida E et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. **2005**;31:1345-1355.
15. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. **2006**;34:1297-1310.
16. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med. **2007**;35:827-835.
17. Jha AK, DesRoches CM, Campbell EG et al. Use of electronic health records in U.S. hospitals. N Engl J Med. **2009**;360:1628-1638.
18. Kuzniewicz MW, Vasilevskis EE, Lane R et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest. **2008**;133:1319-1327.
19. Liu V, Turk BJ, Ragins AI, Kipnis P, Escobar GJ. An electronic simplified acute physiology score-based risk adjustment score for critical illness in an integrated healthcare system*. Crit Care Med. **2013**;41:41-48.
20. Render ML, Deddens J, Freyberg R et al. Veterans Affairs intensive care unit risk adjustment model: validation, updating, recalibration. Crit Care Med. **2008**;36:1031-1042.
21. Johnson AE, Kramer AA, Clifford GD. A New Severity of Illness Scale Using a Subset of Acute Physiology, Age, and Chronic Health Evaluation Data Elements Shows Comparable Predictive Accuracy. Crit Care Med. **2013**;41:1711-1718.
22. Graham PL, Cook DA. Prediction of risk of death using 30-day outcome: a practical end point for quality auditing in intensive care. Chest. **2004**;125:1458-1466.

23. Drye EE, Normand SL, Wang Y et al. Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: an observational study with implications for hospital profiling. Ann Intern Med. **2012**;156:19-26.
24. Pouw ME, Peelen LM, Moons KG, Kalkman CJ, Lingsma HF. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. BMJ. **2013**;347:f5913.
25. Partners Healthcare. Research Patient Data Registry (RPDR). **2013**. http://rc.partners.org/rpdr Last accessed: Nov 11 2013.
26. Higgins TL, Kramer AA, Nathanson BH, Copes W, Stark M, Teres D. Prospective validation of the intensive care unit admission Mortality Probability Model (MPM0-III). Crit Care Med. **2009**;37:1619-1623.
27. Healthcare cost and utilization project. Comorbidity Software, version 3.7. **2012**. http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp Last accessed: Nov 1 2013.
28. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. **1998**;36:8-27.
29. van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariable Imputation by Chained Equations in R. Journal of Statistical Software. **2011**;45:1-67.
30. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag; **2001**.
31. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. **2001**;54:774-781.
32. Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken: J. Wiley & Sons; **1987**.
33. CDC Division of Healthcare Quality Promotion. The National Healthcare Safety Network Device-Associated Module: Ventilator-Associated Event Protocol. **2013** Atlanta, USA
34. Polderman KH, Thijs LG, Girbes AR. Interobserver variability in the use of APACHE II scores. Lancet. **1999**;353:380.
35. Bosman RJ, Oudemane van Straaten HM, Zandstra DF. The use of intensive care information systems alters outcome prediction. Intensive Care Med. **1998**;24:953-958.
36. Pine M, Jordan HS, Elixhauser A et al. Enhancement of claims data to improve risk adjustment of hospital mortality. JAMA. **2007**;297:71-76.
37. Tabak YP, Sun X, Derby KG, Kurtz SG, Johannes RS. Development and validation of a disease-specific risk adjustment system using automated clinical data. Health Serv Res. **2010**;45:1815-1835.
38. Afessa B, Keegan MT, Gajic O, Hubmayr RD, Peters SG. The influence of missing components of the Acute Physiology Score of APACHE III on the measurement of ICU performance. Intensive Care Med. **2005**;31:1537-1543.
39. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med. **2007**;35:2052-2056.
40. Keegan MT, Gajic O, Afessa B. Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance. Chest. **2012**;142:851-858.

Part 2

Administrative data for HAI surveillance

8

Accuracy of hospital discharge coding data for the surveillance of drain-related meningitis

Maaike S.M. van Mourik[a], Annet Troelstra[a], Karel G.M. Moons[b], Marc J.M. Bonten[a,b]

[a] *Department of Medical Microbiology*
[b] *Julius Center for Health Sciences and Primary Care*
University Medical Center Utrecht, Utrecht, The Netherlands

Infection Control and Hospital Epidemiology 2013 Vol. 34(4): 433 – 436.
© 2013 by The Society for Healthcare Epidemiology of America.

NARRATIVE ABSTRACT

Surveillance of healthcare-associated infections is labor intensive and complex. Discharge coding is an accessible source of information that may support detection of cases. For drain-related meningitis, however, discharge coding data had low sensitivity (32%) and positive predictive value (35%) and could neither replace nor improve existing complex surveillance systems.

INTRODUCTION

Surveillance of healthcare-associated infections (HAI) and feedback of HAI rates are widely recognized as a key feature of infection prevention programs. Furthermore, reporting of infection rates is increasingly demanded by policy makers, payers and the public[1,2]. However, efficient and reliable quantification of HAI incidence is not straightforward. Traditional surveillance by manual chart review is time-consuming and error-prone, which has driven the search for more efficient methods, such as classification algorithms and more complex prediction models based on culture results and antibiotic use, among other data[3,4]. Alternatively, routinely collected administrative data, such as hospital discharge coding and billing records, are readily available and could be used to identify patients that have developed an infection[5]. However, the accuracy of administrative data for the surveillance of HAI has not been universally demonstrated[6].

Drain-related meningitis (DRM) is an infection occurring frequently in neurosurgical patients receiving external cerebrospinal fluid drainage devices[7]. We sought to investigate the accuracy of administrative data for the surveillance of DRM and, subsequently, the additional value that such data provided to an existing prediction model[4].

METHODS

Study population

This retrospective cohort study included all patients which received an external ventricular or lumbar drain registered in routine surveillance by the hospital hygiene department from 2004 to 2010 (n = 838), with the exception of children (n = 121), patients with pre-existing meningitis (n = 33), patients who died within 24 hours of drain placement (n = 47), patients admitted with a drain placed elsewhere (n = 5) and military personnel or patients with repeated admission more than 30 days after an initial discharge (n = 15)[4].

Reference standard

Routine surveillance through chart review by infection preventionists was the reference standard. For all patients, the presence of DRM was assessed using a modified Centers for Disease Control and Prevention definition for healthcare-associated meningitis, with adjudication through review in case of disagreement[4]. Administrative data were not taken into account during the chart review.

Prediction model

The multivariable prediction model has been described in detail previously[4]. Data were collected from a clinical data warehouse; use of anonymized data from this data warehouse has been exempted from review by the institutional review board[4]. For each patient, this model returns a predicted probability of DRM to which a threshold can be applied to identify those patients at high risk of having developed DRM. Manual confirmation of meningitis can then be limited to high-risk patients. Selection of model threshold is driven by the user's preferences with regard to sensitivity and specificity.

Administrative data

For all patients, *International Classification of Disease* (ICD-9CM) codes (discharge codes) were extracted from administrative databases for the index admission along with billing records filed during that period. Discharge codes were assigned by medical coders based on discharge summaries and transferred to the national medical registration (LMR). A priori, two investigators (AT and MvM) selected codes that could potentially reflect drain-related meningitis[8]. Because there was no dedicated code for DRM, a range of proxy indicators was selected aiming at maximizing sensitivity while maintaining specificity (**table 1**). To ensure maximal sensitivity, diagnosis-treatment combinations from billing records were subsequently analyzed as well.

Analysis

The sensitivity, specificity and predictive values of discharge codes in comparison to manual surveillance (reference) were assessed. Sensitivity was considered particularly important.The complementary value of administrative data next to the existing multivariable model was assessed in two ways. First, patients were reclassified to the high risk group if they were flagged by a positive discharge code irrespective of their model classification; subsequent changes in sensitivity (number of missed cases) and efficiency (number of false positives that unnecessarily undergo manual review) were examined. Secondly, discharge codes were simply included as an additional predictor within the model and model fit was assessed (likelihood ratio test and change in area under the ROC curve). Analyses were performed using SPSS version 19.

RESULTS

Of the 617 patients included in the analysis, 97 developed DRM according to the reference standard (lumbar drain: 11/208; ventricular drain: 86/409). The mean number

Table 1 | Selected ICD-9 codes and the number of times they were assigned during the index admission for patients with DRM and those without.

Code	Description	Number of times assigned		p-value
		DRM N = 97	No DRM N = 520	
112.83	Candidal meningitis	0	0	--
320.00 – 320.9	Bacterial meningitis	7	4	<0.01
322.00 – 322.9	Meningitis of unspecified cause	12	10	<0.01
324.00 – 324.9	Intracranial and intraspinal abscess	0	2	1.0
349.10	Nervous system complication from surgically implanted device	0	0	--
792.00	Nonspecific abnormal findings in cerebrospinal fluid	0	0	--
996.60	Infection and inflammatory reaction due to unspecified internal prosthetic device, implant, and graft	1	1	0.29
996.63	Infection and inflammatory reaction due to nervous system device, implant, and graft	0	0	--
996.70	Other complication due to unspecified prosthetic device, implant, and graft	0	2	1.0
996.75	Other complication due to nervous system device, implant, and graft	0	0	--
997.00	Nervous system complication, unspecified	11	33	0.12
997.01	Central nervous system complication	0	0	--
997.09	Other nervous system complication	0	0	--
998.50 – 998.59	Post-operative infections	7	3	0.20
999.30 – 999.39	Other infection (excludes listed conditions)	0	1	1.0

Note: P-value by Chi-square or Fisher exact test where appropriate. Abbrviations: DRM – drain-related meningitis, ICD-9 – International Classification of Diseases, 9th revision.

of discharge codes assigned during the admission was 1.88 (range 0 to 8). Discharge codes correctly identified 31 out of 97 cases of DRM (sensitivity 32.0%, 95% confidence interval (CI) 22.9 to 42.2). The specificity was 88.8% (462 out of 520, 95%CI 85.8 to 91.4). Fifty-eight patients were incorrectly flagged as DRM by the discharge codes (positive predictive value 34.8%, 95% CI 25.0 to 45.7) and negative predictive value was 87.5% (95% CI 84.4 to 90.2). Combination with diagnosis-treatment combination codes extracted from the billing records did not meaningfully improve accuracy (sensitivity 34.0%, positive predictive value 34.7%).

Table 2 shows how the number of incorrectly classified patients changed when using discharge codes as an addition to classification by the original prediction model. The number of missed cases only modestly decreased at the higher (user-defined) model thresholds; this decrease, however, came at the cost of additional false positives which require manual chart review (see example in table 2). Adding discharge codes to the

Table 2 | Reclassification based on discharge codes besides the original surveillance system for drain-related meningitis. For several possible model thresholds, changes in numbers of missed infections and false positive signals are shown. Gain in sensitivity is offset by increased numbers of false positives requiring manual chart review.

	Original multivariable model		Model + discharge code reclassification	
P(DRM) threshold	**Missed cases (below threshold)**	**False positives (above threshold)**	**Missed cases (change)**	**False positives (change)**
0.05	0	113	0 (=)	148 (+ 35)
0.10	1	74	1 (=)	118 (+44)
0.15	1	64	1 (=)	110 (+46)
0.20	6	52	5 (-1)	100 (+ 48)
0.25	11	42	9 (-2)	91 (+ 49)
0.30	16	30	13 (-3)	81 (+51)

Note: Example: if a threshold of 0.2 was selected, all patients with a predicted probability of 0.2 or higher are classified as high-risk; the model then has 93.8% sensitivity (91/97) and 64% positive predictive value (91/143). Reclassification by discharge codes slightly improved sensitivity to 94.8%, however positive predictive value decreased to 49.2%.
Abbreviations: P(DRM) – predicted probability of drain-related meningitis.

existing model did not improve model performance (p-value 0.39, area under the ROC curve 0.97 in both models).

DISCUSSION

This analysis shows that administrative data cannot be used as an independent surveillance system for drain-related meningitis. In particular, sensitivity of the discharge codes was very low (32%). The low positive predictive value of 35% may (in part) be explained by the broad range of codes selected as potentially indicative of DRM. In addition, discharge codes have no added value besides the more complex prediction model developed previously; the marginal gain in sensitivity is offset by a loss of efficiency.

These findings are in line with results from other studies showing limited sensitivity of discharge codes for the detection of other HAI[6,9,10]. Since discharge codes are often assigned for purposes other than infection surveillance, the aim may not be to capture all complications that occurred. Furthermore, for certain HAI there are no dedicated codes available, thus complicating their use as a surveillance system. In our setting, codes are mainly assigned based on physician discharge summaries, and therefore also depend on adequate documentation. An exploratory analysis showed that missed infections may be attributed to lack of specific codes, no mention of DRM in discharge summaries, untimely documentation and unexplained missed opportunities for coding.

This study has several limitations. Codes included are only those assigned during the index admissions and not those assigned for subsequent readmissions. However, cases of DRM are virtually always captured during the index admission; most patients remain in the hospital for several days after removal of their drainage device and meningitis must occur within seven days of drain removal to be linked to the device. Furthermore, although this study is based on the ICD-9 system, we do not expect greatly improved performance under the recently implemented ICD-10 system as codes are still assigned in similar fashion and no codes specific for DRM have been added. Finally, for certain HAI specific codes are available and coding practices may differ across healthcare systems, thus results may not be generalizable to other HAI or healthcare systems.

In conclusion, administrative data do not accurately identify cases of drain-related meningitis and caution is necessary when using administrative data for the surveillance of healthcare-associated infections. The validity of discharge coding-based surveillance should be assessed when considering implementation in new settings or for new conditions.

Acknowledgements

The authors would like to thank J.H. den Breeijen, L. Beks and the department of hospital hygiene for their assistance in data collection.

Funding

MvM was funded by an Alexandre Suermann Stipendium from the UMC Utrecht. MB was supported by the Netherlands Organization for Scientific Research (NWO VICI Grant 918.76.611). Conflicts of interest - All authors report no conflicts of interest relevant to this article.

REFERENCE LIST

1. Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med.* **2007**;357:1573-1575.
2. Gaynes R, Richards C, Edwards J et al. Feeding back surveillance data to prevent hospital-acquired infections. *Emerg Infect Dis.* **2001**;7:295-298.
3. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. *Clin Infect Dis.* **2009**;48:1268-1275.
4. van Mourik MS, Moons KG, van Solinge WW et al. Automated Detection of Healthcare Associated Infections: External Validation and Updating of a Model for Surveillance of Drain-Related Meningitis. *PLoS ONE.* **2012**;7:e51509.
5. Calderwood MS, Ma A, Khan YM et al. Use of Medicare Diagnosis and Procedure Codes to Improve Detection of Surgical Site Infections following Hip Arthroplasty, Knee Arthroplasty, and Vascular Surgery. *Infect Control Hosp Epidemiol.* **2012**;33:40-49.
6. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. *Clin Infect Dis.* **2009**;49:949-955.
7. Lozier AP, Sciacca RR, Romagnoli MF, Connolly ES, Jr. Ventriculostomy-related infections: a critical review of the literature. *Neurosurgery.* **2002**;51:170-181.
8. Centers for Disease Control and Prevention. International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9CM), Sixth Edition. **2010**. http://www.cdc.gov/nchs/icd/icd9cm.htm#ftp Last accessed: Sep 5 2011.
9. Moro ML, Morsillo F. Can hospital discharge diagnoses be used for surveillance of surgical-site infections? *J Hosp Infect.* **2004**;56:239-241.
10. Curtis M, Graves N, Birrell F et al. A comparison of competing methods for the detection of surgical-site infections in patients undergoing total arthroplasty of the knee, partial and total arthroplasty of hip and femoral or similar vascular bypass. *J Hosp Infect.* **2004**;57:189-193.

9

Accuracy of administrative data for surveillance of healthcare-associated infections: a systematic review

Maaike S.M. van Mourik[a], Pleun Joppe van Duijn[b], Grace M. Lee[c], Karel G.M. Moons[b], Marc J.M. Bonten[a,b]

[a] *Department of Medical Microbiology, University Medical Center Utrecht, The Netherlands*
[b] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands*
[c] *Department of Population Medicine, Harvard Pilgrim Healthcare Institute, Harvard Medical School, Boston, Massachusetts, USA*

Manuscript in preparation

ABSTRACT

Background – Measuring the incidence of healthcare-associated infections (HAI) is of increasing importance in current healthcare delivery systems. Administrative data, including discharge diagnoses, are commonly used to determine the occurrence of HAI, either as quality indicators or to support within-hospital surveillance programs.

Objective – To evaluate the accuracy of administrative data for the detection of HAI, both for within-hospital surveillance and for external quality assessment.

Methods – Systematic review of studies assessing the accuracy of HAI surveillance using administrative data (1995-March 2013). Rigorous methodological quality assessment was performed using QUADAS-2 criteria; accuracy estimates were stratified by HAI type and key study characteristics.

Results – 57 studies evaluated the accuracy of administrative data for HAI surveillance. The majority targeted surgical site or bloodstream infections. Study designs were very diverse with respect to the selection of administrative data, the outcome definitions used to determine HAI presence and the extent to which the ascertainment of HAI was complete. Observed sensitivity and positive predictive values were very heterogeneous and generally modest at best for all HAI types. Identification of device-associated infections was particularly problematic. No apparent differences were found between algorithms using administrative data for within-hospital surveillance systems or as quality indicators. The large heterogeneity in design across studies precluded formal calculation of summary accuracy estimates.

Conclusion – Administrative data had limited, and highly variable, accuracy for the detection of HAI, and their judicious use for both internal surveillance efforts and external quality assessment is recommended. Future, robust, validation studies to properly quantify the diagnostic accuracy of administrative data are imperative if these continue to be used for HAI surveillance.

INTRODUCTION

Assessment of quality of care and monitoring of patient complications is a key concept in current healthcare delivery systems[1]. Administrative data, and discharge codes in particular, have been used as a valuable source of information to define patient populations, assess severity of disease, determine patient outcomes and detect adverse events, including healthcare-associated infections (HAI)[2-5]. In certain instances, administrative data are employed to measure quality-of-care and derive payment incentives or govern nonpayment for (infectious) complications. Examples include the patient-safety indicators (PSIs) developed by the Agency for Healthcare Quality Research, nonpayment for Healthcare-Associated Conditions (HAC) determined to be preventable 'never-events' and the expansion of value-based purchasing by the United States (U.S.) Centers for Medicare and Medicaid Services[6-9]. Although clinical patients outcomes such as HAI rates reported to the National Healthcare Safety Network (NHSN) are increasingly used in quality assessment, administrative data are still a key component of quality evaluations[4,7,10].

Within the process of HAI surveillance, administrative data can have several targeted applications. Firstly, they may be used to support within-hospital surveillance efforts, for example by flagging patients that probably developed a HAI and require further review. For this purpose, administrative data can either be used in isolation or be combined with other indicators of HAI such as microbiology culture results or antibiotic dispensing. Alternatively, discharge codes may be used in (external) quality indicator algorithms that directly determine the occurrence of HAI and thus gauge hospital performance, compare hospitals or drive payment incentives[3,11,12]. The targeted application of administrative data will determine what measures of concordance are most important to determine the accuracy of the algorithm. When used as an (adjunct) case-finder for within-hospital surveillance, sensitivity may be preferred over positive predictive value to identify medical records requiring manual review. Conversely, in the setting of external quality assessment and comparison across hospitals, high positive predictive value of observed signals may be of greater importance than detecting all cases of HAI.

Nonetheless, many cautionary notes have been raised regarding the accuracy of administrative data for the purpose of HAI surveillance[1,12-14]. Their universal use, ease of accessibility and relative standardization across settings and time makes them attractive for large-scale surveillance and research efforts. On the flip side, inherent to their purpose as means to organize billing and reimbursement of healthcare, administrative data were not designed for the surveillance of HAI. When assigning primary and secondary discharge diagnosis codes, many other interests have greater priority - including financial

incentives – and the reliability of diagnosis code assignment depends heavily on adequate clinician documentation and the number of diagnoses in relation to the number of slots available[3,15].

The primary objective of this systematic review was to assess the accuracy of administrative data for the surveillance of a broad range of HAI, and to determine whether accuracy of algorithms developed for within-hospital surveillance efforts differs from those meant for external quality evaluation. In addition, we rigorously evaluated the methodological quality of included studies, assessed the impact of possible risk of bias and explored other determinants of accuracy.

METHODS

Studies evaluating the accuracy of administrative data in detecting HAI were included in this systematic review, with the exception of studies specifically assessing infections with particular pathogens (e.g. methicillin-resistant *Staphylococcus aureus* or *Clostridium difficile*). Manual review of patient clinical records to assess the presence of HAI was considered the reference standard and the index test was defined as selection(s) of discharge and/or procedure codes extracted from administrative data. The results of this analysis are reported in accordance with the guidelines of the PRISMA statement (Preferred Reporting Items for Systematic Reviews and Meta-Analysis)[16]. This review did not receive protocol registration.

Search

Medline, EMBASE, the Cochrane database and CINAHL were searched for studies published starting 1995 with a query combining representations of administrative data and (healthcare-associated) infections (supplementary data 1 (**table S1**)). Limits were set to articles published in English, French or Dutch and the search was performed on March 8th 2012. The search was closed on March 1st 2013.

Study selection

To define suitability for inclusion, the following criteria were applied: 1) the study assesses the concordance between administrative data and HAI and includes some type of reference standard for HAI, 2) data was collected (in part) in 1995 or later, 3) the study does not reflect syndromic surveillance or employs only natural language processing and 4) the study presents original research. Selection of studies was done by a single reviewer (MvM). All full-text articles were cross-referenced to detect possibly missed studies. The

study selection was not restricted to specific geographical locations or patient populations nor was there a requirement for complete data availability.

Quality assessment and data extraction

After selection of studies, quality assessment and data-extraction was performed by two independent reviewers (MvM, PJvD) using modified QUADAS-2 criteria for quality assessment of diagnostic accuracy studies[17,18] (see **table S2** for details, assumptions and simplifications made). In brief, the QUADAS-2 criteria evaluate risk of bias and concerns with applicability to the research question with respect to methods of patient selection, the index test (administrative data algorithm) and the reference standard (method of HAI ascertainment). In addition, the criteria provide a framework to evaluate risk of bias introduced by the (in)completeness of HAI ascertainment, so-called 'patient flow'. Agreement was reached by discussion.

Several studies contained multiple study designs; quality assessment and data-extraction was then applied separately to each comparison. Special attention was given to blinding of abstractors when determining HAI status as well as to partial or differential verification patterns. Partial verification occurs when not all patients are assessed for HAI development, in a pattern that relies on the result of the index test (i.e. administrative data). In the case of differential verification, not all patients receive the same reference standard conditional on the result of the administrative data. Depending on the pattern of partial and/or differential verification, this may introduce bias in the observed accuracy estimates, as outlined in *Naaktgeboren et al*[19].

Analyses

Included studies were stratified by HAI type, risk of bias based on QUADAS-2 criteria and the purpose of the administrative data within HAI surveillance. Studies were grouped as meant for internal within-hospital, surveillance either as the sole method of case-finding or combined with other HAI indicators, or as studies assessing a quality indicator for external use (e.g. public reporting or pay-for-performance programs). Forest plots were created for sensitivity, specificity, positive and negative predictive values where possible. If large enough groups of sufficiently comparable studies with complete two-by-two tables were available, accuracy estimates were pooled using the bivariate method recommended in the Cochrane Handbook for Systematic Reviews of Diagnostic Accuracy[20,21]. This analysis jointly models the distribution of sensitivity and specificity, accounting for correlation between these two outcome measures. All analyses were performed using R version 3.0.1 (www.r-project.org) and SPSS Statistics 20 (IBM, Armonk, NY).

RESULTS

Study selection

After removal of duplicate references, 8478 unique titles were screened for relevance to the research question and exclusion criteria were applied to the 675 remaining abstracts. Four additional articles were found through cross-referencing and 10 relevant articles were identified between the search date and search closure (**figure 1**). Fifty-seven studies, containing 71 comparisons, were available for the qualitative synthesis and underwent methodological quality assessment[22-78].

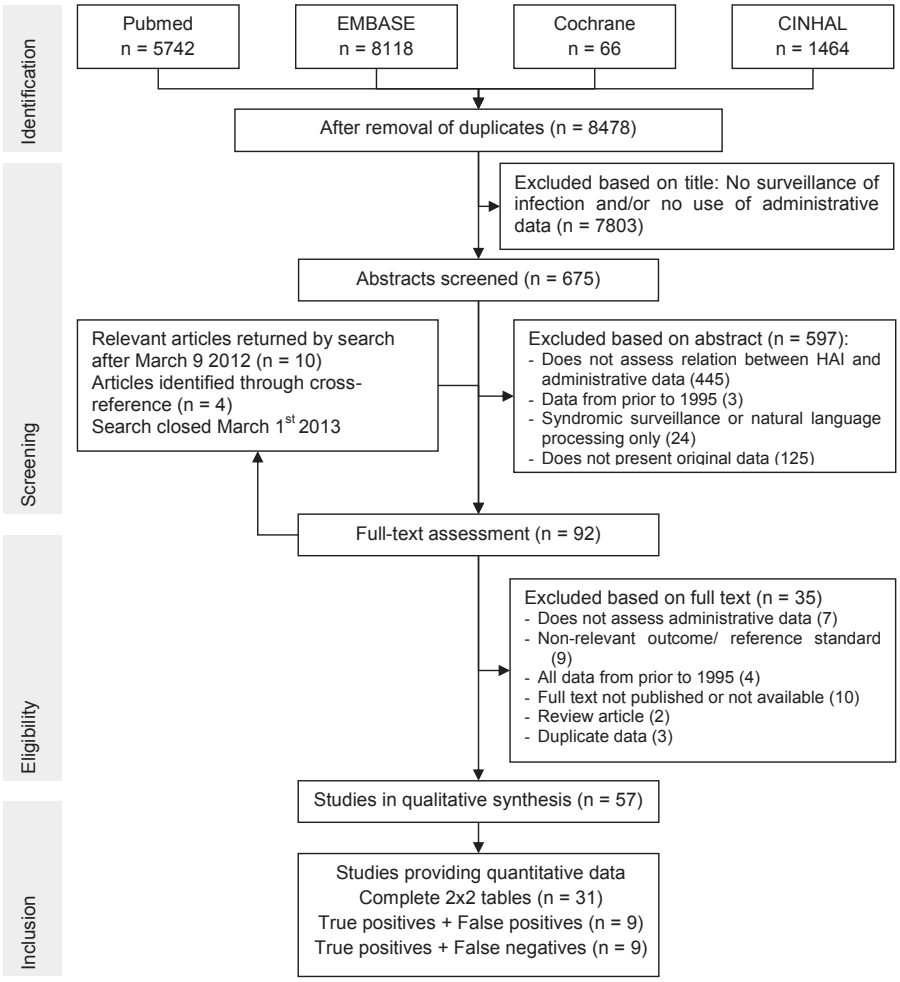


Figure 1 | Flowchart of study selection and inclusion.

Study characteristics

The included studies varied greatly with respect to study design, methodology used as reference standard and the selection of discharge codes included as the index test. Thirty-five studies applied National Healthcare Safety Network (NHSN) methodology (or a translation thereof) to determine HAI presence, six defined HAI as registered in the (National or Veteran's Administration) Surgical Quality Improvement Project (NSQIP, VASQIP), and the remaining studies used clinical (4) or other (12) methods (**table 1**). Case-definitions were applied by infection preventionists in 24 studies, but also by trained nurses (9), physicians (9) or other abstractors. Thirty-three studies assessed algorithms for within-hospital (internal) surveillance efforts; in fifteen, administrative data were combined with other indicators of infection to detect HAI (e.g. microbiology culture results or antibiotic use). Twenty-four studies assessed administrative data algorithms explicitly designed for external use. Only a few studies provided data collected after 2008[32,35,37,46,54,67,70]. The majority of studies were conducted in the U.S. (n = 44) and 8 originated from European countries.

Methodological quality

Methodological quality assessment revealed that ascertainment of HAI status was complete in 37 of 57 studies; in other words, in only 65% of studies the same definition and method HAI detection was applied to (a random sample of) all patients. Alternative verification patterns were evaluation of patients flagged by administrative data only (9 studies), assessment of patients flagged by either administrative data or another test (e.g. a positive culture result) (8) and reclassification of discrepant cases after a second review. Also blinding of the assessors to the result of administrative data was not always done or not adequately described. **Figure 2** summarizes the risk of bias and applicability concerns for each QUADAS-2 domain (table **S3** for details by study; **figure S1** for summary figures by HAI type). A high risk of bias for the flow component was particularly observed in studies that did not allow for extraction of complete two-by-two tables, mainly as a result of partial verification. Of note, in studies assessing only the positive predictive value, only patients flagged by administrative data undergo HAI ascertainment. This partial verification in itself does not introduce bias but lack of blinding of assessors may still result in an overall risk of bias.

Table 1 | Main characteristics of included studies. Some study presented multiple comparisons and/or assessed more than 1 type of healthcare-associated infection.

	Total	SSI	BSI	UTI	Pneum	Other
N studies	**57**	**34**	**24**	**15**	**14**	**2**
(N comparisons)	**(71)**	**(44)**	**(29)**	**(15)**	**(15)**	**(2)**
Device-associated	20	--	12	7	9	1
ICU only	5	1	3	2	3	0
Type of reference standard						
-NHSN	35	26	9	6	7	2
-(VA)SQIP	6	2	6	2	3	0
-Clinical	4	1	3	1	1	0
-Other	12	5	6	6	3	0
Application of administrative data						
-External quality assessment	24	9	19*	6	8	0
-Within hospital surveillance	18	13	3	7	4	1
-Combined with other HAI indicators	15	12	3	2	2	1
Specific quality metric						
-PSI	9	1	10	0	2	0
-HAC	3	0	2	1	0	0
-PHC4	4	4	3	3	4	0
Region of origin						
-United States	44 (55)	22 (29)	19 (24)	10 (10)	9 (10)	1 (1)
-Europe	8 (10)	8 (9)	4 (4)	4 (4)	4 (4)	1 (1)
-Other	4 (6)	4 (6)	1 (1)	1 (1)	1 (1)	0 (0)
High risk of bias on QUADAS domain						
-Patient selection	1 (1)	1 (1)	1 (1)	0 (0)	1 (1)	0 (0)
-Index test	0 (3)	0 (1)	0 (1)	0 (0)	0 (0)	0 (0)
-Reference standard	19 (27)	11 (18)	6 (7)	4 (4)	2 (2)	1 (1)
-Flow	19 (29)	10 (18)	8 (11)	4 (4)	3 (4)	1 (1)
Verification pattern						
-Complete or random sample	37 (42)	23 (26)	16 (18)	11 (11)	10 (10)	1 (1)
-Complete with discrepant analysis	3 (6)	3 (6)	1 (2)	1 (1)	1 (2)	0 (0)
-Partial, based on index test only	8 (8)	2 (4)	5 (7)	2 (2)	2 (2)	0 (0)
-Partial, based on index and other test	8 (12)	6 (6)	1 (1)	1 (1)	1 (1)	1 (1)
-Other or unclear	1 (3)	0 (2)	1 (1)	0 (0)	0 (0)	0 (0)
Data availability						
-Complete 2x2 table, by HAI type	29	20	10	6	6	1
-Complete 2x2 table, HAI combined	3	3	2	4	3	0
-Positive predictive value only, by HAI	9	3	6	1	2	0
-Other	9	2	5	3	3	0
-No data extraction possible	7	6	1	1	0	1

*one study targeting external quality assessment using administrative data combined with other sources of data.

Abbreviations: HAC – Healthcare-associated condition as defined by the Centers for Medicare and Medicaid Services, ICU – intensive care unit, NHSN – National Healthcare Safety Network, PSI – Patient Safety Indicator, PHC4 – Pennsylvania Healthcare Cost Containment Counsel code selection, (VA)SQIP – (Veteran's Administration) Surgical Quality Improvement Project, QUADAS – Quality assessment for diagnostic accuracy studies.

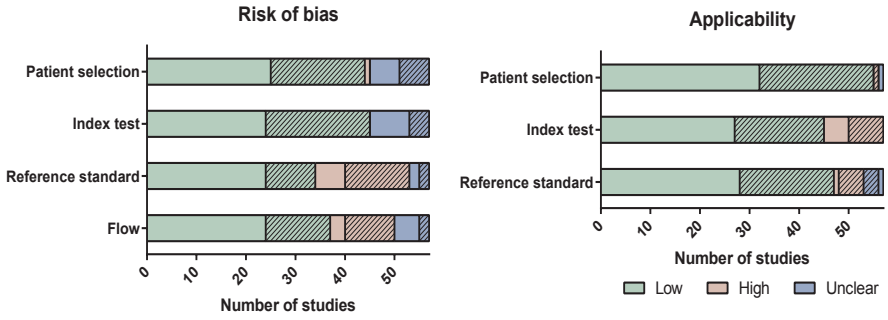


Figure 2 | Summary of risk of bias and applicability for all studies (n = 57), assessed using the Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2) methods.

Some studies contain multiple comparisons; in this case the lowest risk of bias per study is included. Shading denotes studies where extraction of complete two-by-two tables was not possible, including studies only assessing positive predictive values.

Surgical site infection (SSI)

Thirty-four studies assessed SSI. Most studies identified patients at risk of SSI by identifying procedure codes from claims data, although a few included all patients admitted to surgical wards. Studies applying NHSN surveillance methods generally also incorporated diagnosis codes assigned during readmissions in their algorithm to complete the required follow-up duration, and several algorithms also included follow-up procedures to detect SSI. Twenty studies allowed for extraction of two-by-two tables and an additional three investigated only positive predictive value. **Figure 3A** presents the forest plots of sensitivity and positive predictive value for the studies providing quantitative data (specificity and negative predictive value in **figure S2A**).

The accuracy estimates were highly variable, also within groups of studies with similar intended application and target procedures. Given the small number of studies within each subgroup, no pooling of studies was attempted. Several studies assessed different selections of administrative data within one study; as would be expected, using a broader selection of discharge codes detects more cases of SSI at the cost of lower positive predictive value[27,48,55]. Between the studies, this relation between the specificity of the discharge codes included and the observed accuracy measures was not apparent (998.5 or 996.6 vs. a broader selection, data not shown). Visual inspection of the forest plots suggests that – in general – study designs with a higher risk of bias had more favorable performance characteristics than those with more robust methodological quality.

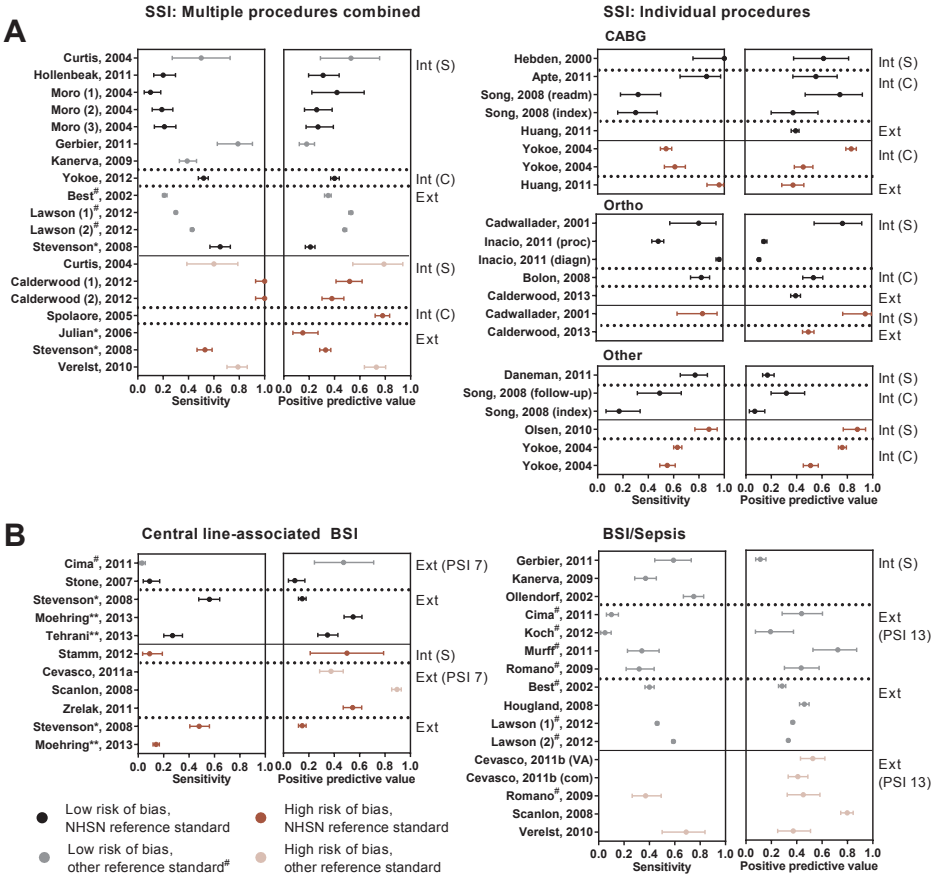
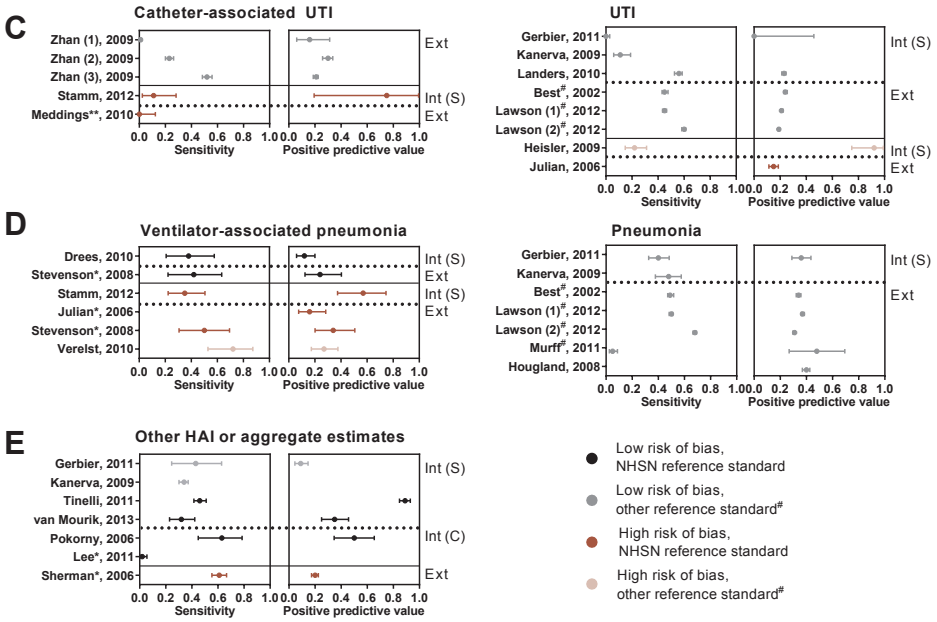


Figure 3 | Forest plots for sensitivity and positive predictive value, stratified by HAI type and relevant study characteristics.

Bloodstream infections (BSI)

Of the 24 studies evaluating bloodstream infections, half focused on central line-associated BSI (CLABSI), 19 assessed algorithms for external quality assessment. Methods of determining central line utilization for CLABSI were very diverse; in studies evaluating PSI 7 ('central venous catheter-related BSI') or HAC this was determined by specific discharge codes, other studies only included patients with positive blood cultures[68] or used traditional surveillance data to determine the population at risk[70]. The sensitivity of CLABSI detection was no higher than 40 percent in most cases. Notably, only the studies that did not rely on administrative data to determine central line presence achieved sensitivity over 20% (**figures 3B** and **S2B**). In general terms, the sensitivity for

Figure 3 | *Continued*



Studies are grouped by the intended application of administrative data:
Int (S) – used in isolation to support within-hospital surveillance efforts.
Int (C) – used to support within-hospital surveillance, combined with other indicators of infection.
Ext – used for external quality assessment, including public reporting and pay-for-performance.
Notes: In studies including multiple specifications of the administrative data algorithm, these are numbered sequentially. 95% confidence intervals are derived using the exact binomial method. If multiple study designs were performed within a single study, they are mentioned separately.
#: reference standard from Surgical Quality Improvement Project (NSQIP or VASQIP). *: Code selection based on specification from Pennsylvania Health Cost Containment Council. **: HAC specification.
Abbreviations: BSI – bloodstream infection, CABG – coronary artery bypass graft, DRM – drain-related meningitis, Ortho – orthopedic procedure, PSI – Patient safety indicator, Sep – Sepsis, SSI – surgical site infection, UTI – urinary tract infection.

detecting BSI was slightly better. The large variability in study characteristics precluded the generation of summary estimates in most subgroups. The pooled sensitivity of PSI 13 ('post-operative sepsis') in studies using SQIP methods as a reference standard was 17.0% (95% confidence interval 6.8-36.4) with a specificity of 99.6% (99.3-99.7). Of the algorithms targeted at external quality assessment, the positive predictive values were inconsistent and mediocre, suggesting these quality indicators detected many events that were not (CLA)BSI. As for SSI, study designs with higher risks of bias tended to show higher accuracy, in particular for positive predictive value.

Urinary tract infection (UTI)

Of the fifteen studies investigating urinary tract infection, 7 focused specifically on catheter-associated UTI (CAUTI). When the administrative data algorithm also incorporated identification of the patients receiving a urinary catheter[79,80], the low sensitivity of CAUTI detection is striking (**figure 3C**, **S2C**). The study by *Zhan et al.*[80] demonstrates that inadequate coding of catheterization leads to poor sensitivity of CAUTI detection and how this can be improved somewhat by expanding the denominator to all surgical patients. Sensitivity was higher for UTI, but positive predictive values were universally below 25% except in the study by *Heisler et al.*[39]. This study additionally scrutinized flagged records to exclude UTI.

Pneumonia

Fourteen studies evaluated pneumonia, of which 9 targeted ventilator-associated pneumonia (VAP). The presence of mechanical ventilation was either determined within the administrative data algorithm[35,44] or determined by manual methods[68]. In the studies assessing VAP that allowed for calculation of accuracy measures, sensitivity ranged from 35 to 72% and positive predictive value from 12 to 57%. Pooling of studies was not undertaken given the small number of studies in each group. For pneumonia, sensitivity and positive predictive value hovered around 40% although it must be emphasized that the studies used very diverse methodologies (**figure 3D, S2D**).

Algorithms combining administrative data with clinical data

Fifteen studies in this review evaluated the accuracy of administrative data in a combined approach with other (automated) indicators of HAI for within-hospital surveillance. Ten allowed for extraction of accuracy estimates of administrative data alone and only very few provided the data necessary to fairly assess the incremental benefit of administrative data to clinical data such as antimicrobial dispensing or microbiology results. In these studies the gain in sensitivity obtained by adding administrative data was generally modest (at most 10 percent points)[24,50,51,60,75,76].

DISCUSSION

In light of the increasing attention for evaluating, improving and rewarding quality of care, efficient and reliable measures to detect HAI are vital. Administrative data are commonly used to monitor HAI incidence, however as demonstrated by this comprehensive systematic review, they have limited – and very variable – accuracy for the detection of HAI. There was no clear difference in accuracy of administrative data algorithms developed for the purpose of supporting within-hospital surveillance versus those meant

for external quality assessment. Sensitivity was highly variable for both purposes and also algorithms targeting very specific events (CAUTI, CLABSI) for external benchmarking and payment rules had positive predictive values that were modest at best in most studies. In addition, identification of invasive devices such as central lines and urinary catheters was problematic using administrative data; these algorithm had lower sensitivities than those where linkage to procedure codes was more straightforward (e.g. SSI) or not necessary to determine the population at risk. The included studies were very heterogeneous in their design, study location, specification of administrative data algorithms and definition and methods of assessment of HAI presence. Thorough methodological quality assessment revealed that incomplete ascertainment of HAI status and/or lack of blinding of assessors was prevalent, thus introducing risk of bias and complicating the balanced interpretation of accuracy estimates.

The drawbacks of administrative data for the purpose of HAI surveillance have been emphasized previously, especially from the perspective of (external) interfacility comparison or the use of administrative data as the sole case finder for within-hospital surveillance. The added value of administrative data to enhance case-finding for routine within-hospital surveillance or within algorithmic automated surveillance systems combining several sources of electronic data is still advocated, however[3,12,14,81]. In comparison with a prior review, we identified a larger number of primary studies (in part due to broader inclusion criteria), distinguished between administrative data algorithms developed for different purposes and rigorously assessed methodological quality. We confirmed the limited accuracy of administrative data previously described, also for specific quality metrics. In addition, in the few studies that allowed assessment of the incremental value of adding administrative data to (automated) algorithmic surveillance systems using other sources of electronic routine care data, gains in sensitivity were modest. There was no apparent relation between the specificity of the algorithm employed and the observed accuracy estimates across studies, although differences may also have been masked by the heterogeneity of the studies. Several studies within the domain of this systematic review have been published since closure of the search, with findings in line with the observations in this review[82-88].

A number of studies explored reasons for the inaccuracy of administrative data for the purpose of HAI detection. For specific quality measures, differences in HAI definitions may account for a portion of the discordant cases[89]; other explanations include the erroneous detection of infections present-on-admission or infections not related to the targeted device, incorrect coding, poor clinician documentation or the limited number of coding slots available in addition to the challenges with identifying invasive

devices[45,52,54,70,77,86,87]. The precarious balance between the accuracy of administrative data and their use in quality measurement and pay-for-performance programs has been argued previously, especially as these programs may create incentives that further undermine the accuracy of administrative data[14]. Recent studies have provided mixed evidence regarding a change in coding practice in response to introduction of pay-for-performance or public reporting programs[90-92].

Several refinements of administrative data are currently in progress that may improve the future accuracy of administrative data for HAI detection. Introduction of presence-on-admission (PoA) indicators to coding systems is expected to improve the accuracy of detection. Only very few studies included in this review explicitly incorporated PoA indicators in their algorithm and although PoA indicators are increasingly available, their actual adoption and accuracy in the process of code assignment remains to be validated[79,93]. In addition, the ongoing to transition from the 9th revision of the *International Classification of Disease* to the 10th revision (ICD-10) may provide increased specificity due to the large number of codes and thus greater granularity[94]. This review could not assess this effect as only seven studies used the ICD-10, often in a setting that was not directly comparable to settings using the ICD-9 (e.g. the U.S.), and some studies purposefully mapped the ICD-10 codes to mimic the ICD-9[33,49]. In recent years, the number of coding slots available in (standardized) billing records has gradually increased, allowing for more secondary diagnoses to be recorded; whether this expansion beyond 15 slots will benefit the registration of HAI and other complications is unclear[61,95]. Finally, this systematic review could not provide sufficient data to evaluate changes in coding accuracy since the 2008 introduction of the discharge diagnosis-based payment incentives in the United States. Ongoing studies are needed to assess the impact of these changes in coding systems on their accuracy for HAI surveillance.

Based on the available data it seems that studies employing designs associated with higher risk of bias provide a more optimistic picture of administrative data accuracy than more robust methodologies. Conversely, administrative data did detect additional cases of HAI that were initially overlooked by manual surveillance in several studies, demonstrating the well-known limitations with quality of traditional surveillance[24,26,27,33,64,68,75]. These concerns with the reference standard, in conjunction with the workload of applying manual surveillance to large numbers of patients, are possible explanations for the frequent use of study designs that include partial or differential verification patterns. Although reclassifying missed cases after a second review will result in more accurate detection of HAI, the differential application of the second review may bias the performance estimates upwards[19]. This is also supported by the higher accuracy observed in studies suffering

from this partial or differential HAI ascertainment. Re-review of a random sample of patients records that were not detected by either manual surveillance or administrative data revealed additional missed cases of HAI in several studies[24,68], and ideally this second review should also be applied to the concordant HAI-positive cases[83]. Future study designs could be improved in several ways, in the first place by ensuring application of the same reference standard to all patients, or - if unfeasible - to a random sample in each subgroup of the two-by-two table. Also, to facilitate a balanced interpretation of the results, accuracy observed before and after reclassification should be reported seperately[96]. Finally, (reporting of the) blinding of reviewers determining HAI status to the results of the administrative data may also lead to more robust performance estimates.

The current systematic review aimed to provide a comprehensive overview of the existing evidence regarding the accuracy of administrative data for HAI detection. Despite efforts to identify all available studies, we cannot exclude the possibility of having missed studies nor did we assess publication bias. In addition, as a result of our broad inclusion criteria, the included studies were very diverse, complicating interpretation of the results. Contrary to previous a systematic review[12], the small number of comparable studies motivated us to refrain from generating pooled summary estimates for most subgroups of HAI. Finally, many of the recent changes in administrative data generation and coding systems were not yet evaluated in the studies included in this review and ongoing research will be necessary to assess their effect on coding accuracy.

CONCLUSION

Administrative data have limited, and highly variable, accuracy for the surveillance of HAI. Sensitivity was insufficient in most studies and also those algorithms targeting specific adverse events for external quality reporting had generally poor positive predictive value. Identification of device-associated infections appeared particularly difficult. The relative paucity of robust studies and the diverse nature of the studies in terms of study designs and methodological quality, together with ongoing refinements in coding systems precludes reliable forecasting of the accuracy of administrative data in future applications. Ongoing and robust validation studies are paramount if the use of administrative data for the purpose of HAI surveillance is to be continued, in particular if administrative data are the sole contributor to measuring patient outcomes.

SUPPLEMENTARY DATA

Table S1 | Search Strategy.

Databases: Medline/Pubmed, EMBASE, CINAHL, Cochrane. All searches in Titles + Abstract
Limits: Published between after 1995, Languages: English, Dutch, French, German. Search dates: Initial search march 8th 2012, search closure March 1st 2013.

Outcome: Healthcare associated infection	Search terms : Infection, infections, hai, infectious, sepsis, meningitis, notifiable, SSI, VAP, pneumonia, CAUTI, CLABSI, CABSI, BSI
AND	
Determinant: administrative data	Search terms : ICD, international Classification of Diseases, administrative, discharge diagnos*, registry, registries, electronic data, claim data, claims data, reimbursement, health plan data, healthplan, medicare, diagnostic coding, discharge coding, discharge code(s), diagnostic coding, diagnostic code(s), diagnosis code(s), diagnosis coding, procedure code(s), procedure coding

Table S2 | Data collection, quality assessment items and assumptions.

GENERAL CHARACTERISTICS		
Item	Options	Considerations & assumptions
Author, year of publication		
HAI studied	SSI/BSI/sepsis/ CLABSI/VAP /UTI/CAUTI/Other	More than 1 may apply Specify details
Systematic post-discharge surveillance?	Yes/No	Only code as yes if explicit aim of the study.
Location of study	Country	
Number of participating centers		
Start and stop of patient inclusion		
Validation of previously developed algorithm	Yes/No	E.g. previous study, PHC4, PSI, HAC
Validation sample within the study	Yes/No	
Purpose of administrative data	Billing/ benchmarking /demographic/ unclear	If U.S.: code as billing
Setting: Medicare, VA or HMO only?	Yes/No (specify)	
Healthcare setting	Primary care, Inpatient, Outpatient, ICU	More than 1 possible
Academic hospital	Yes/No/Mixed (if multicenter)	
Public reporting	Yes/Potentially/No	Was the measure developed/tested as a means of public reporting or external quality benchmarking (as opposed to an in-hospital screening algorithm)
ASSESSMENT OF RISK OF BIAS (ADAPTED FROM QUADAS-2)		
PATIENT SELECTION		
1 Method of patient selection	Describe in-/exclusion criteria	
2 Consecutive or random sample of patients enrolled	Yes/no	Random sampling scored as yes
3 Case-control design avoided	Yes/No	
4 Inappropriate exclusions avoided?	Yes/No	Is the sample enrolled representative of the domain (e.g. no exclusion of high-risk patients?)
5 Risk of bias patient selection	Low/Unclear/High	If#2, #3 or #4 = no, consider risk of bias
6 Applicability patient selection	Low/Unclear/High	
INDEX TEST		
1 Describe index test	Coding system used? Codes assigned by? Procedure codes to detect HAI? PSI algorithm List codes used, duration of follow-up	ICD-9 or ICD-10 Coders, physicians, other, unclear (US: professional coders assumed) No if only used to identify patients at risk Version number Specify use of pre-defined methods (PHC4, PSI, CMS…).

2	Were other tests assessed	Yes/No, specify	
3	Was the administrative data intended as the sole method of surveillance	Yes/no	E.g. were results of administrative data intended to be combined with microbiology results?
4	Was interpretation done without knowledge of the reference standard?	Yes/no	Were codes assigned without knowledge of reference standard?
5	Pre-specified threshold	Yes/no	Was code selection determined in advance? If unspecified and only a very specific code is used, also code as yes (e.g. 998.5 for SSI)
6	Risk of bias index test	Low/Unclear/High	If #4 or #5 = No, consider risk of bias.
7	Applicability index test	Low/Unclear/High	If #3 = No, score as High
REFERENCE STANDARD			
1	Describe reference standard	Method: Definitions used: Applied by:	Describe NHSN/NNIS, (VA)SQIP, Clinical, Other IP, trained nurses, physicians, other abstractor
2	Is the reference standard likely to correctly classify the patient	Yes/No	
3	Was it interpreted without knowledge of the index test?	Yes/No	If only patients flagged by code are received reference standard and/or coding status was unblinded score as No
4	Risk of bias	Low/Unclear/High	If #3 = No, consider risk of bias
5	Applicability	Low/Unclear/High	
FLOW AND TIMING			
1	Describe patients who did not receive 1 of both tests or are not in 2x2 table		Draw flowchart
2	Did all patients receive the RS?	Yes/No	If only assessing patients with positive reference test, score as No
3	Did all patients receive the same RS?	Yes/No	If all the patients receiving RS do not receive the same RS score as No.
4	Were all patients included in the analysis?	Yes/No	
5	Could the patient flow have introduced bias and why?	Low/Unclear/High	If #2 or #3 = Yes, consider risk of bias. If a large or important portion of patients are excluded (e.g. due to missing data), consider risk of bias.
6	How were missing data handled?	Description	

DATA EXTRACTION			
	HAI present	HAI absent	Total
Codes +	TP	FP	
Codes -	FN	TN	
Total			
IF ONLY OUTCOME MEASURES ARE REPORTED:			
Sensitivity		PPV	
Specificity		NPV	

LR-	LR+	
Kappa	Degree of certainty	High – med – low

Notes: General remarks and assumptions: If multiple index tests and/or reference standards and/or patient flow schemes are used in the study, all are assessed separately for their risk of bias (multiple comparisons). Data were extracted for each comparison presented, and also separately if there were a) multiple types of HAI or b) multiple comparisons for each HAI or c) multiple specifications of administrative data.

Table S3 | Risk of bias individual studies, stratified in case of multiple comparisons.

Author, year	HAI studied	Loc	N Center	Study period	definition	Intend appl	N	Risk of bias Pat sel	Index test	Ref	Flow	Applicability Pat sel	Index test	Ref	RoB PPV
Apte, 2011	SSI,	USA	1	2007	Unclear	Int (C)	2	Low	Low	High	Low	Low	Low	High	High
Apte, 2011	SSI,	USA	1	2007	CDC NHSN	Int (C)		Low	Low	Low	Uncl	Low	Low	Low	Uncl
Best, 2002	SSI, Sepsis, Pneu, UTI,	USA	123	1994-1995	(VA/N)SQIP	Ext	1	Uncl	Low	Low	Low	Low	Low	Low	Uncl
Bolon, 2009	SSI,	USA	8	2002-2005	CDC NHSN	Int (C)	1	Low	Low	Uncl	Low	Low	High	Low	Low
Braun, 2006	BSI,	USA	28	1999	Clinical	Ext*	1	Uncl	Low	High	High	Low	Low	High	High
Cadwallader, 2001	SSI,	AUS	1	1998-1999	CDC NNIS	Int (S)	2	Low	Low	Low	Low	Low	Low	Low	Low
Cadwallader, 2001	SSI,	AUS	1		CDC NNIS	Int (S)		Low	Low	High	High	Low	Low	Low	High
Calderwood, 2012	SSI,	USA	4	2007	CDC NHSN	Int (S)	1	Uncl	Uncl	High	High	Low	Low	Low	High
Calderwood, 2013	SSI,	USA	3296	2005-2007	CDC NHSN	Ext	2	Low	Low	High	PPV	Low	Low	Low	High
Calderwood, 2013	SSI,	USA	3296		CDC NHSN	Ext		Low	High	High	PPV	Low	Low	Low	High
Campbell, 2011	SSI, UTI,	USA	1	2008	Other	Int (S)	1	Uncl	Uncl	Low	Low	Low	Low	High	Uncl
Cevasco, 2011a	CLABSI,	USA	28	2002-2007	Other	Ext PSI 3.1	1	Low	Low	High	PPV	Low	Low	Low	High
Cevasco, 2011b	Sepsis,	USA	75	2003-2007	Other	Ext PSI 3.1	2	Low	Low	High	PPV	Low	Low	Low	High
Cevasco, 2011b	Sepsis,	USA	75		Unclear	Ext PSI 3.1		Low	Low	High	PPV	Low	Low	Low	High
Cima, 2011	CLABSI, Sepsis,	USA	1	2006-2009	(VA/N)SQIP	Ext PSI 3.1	1	Low	Low	Low	Low	Low	Low	Low	Low
Curtis, 2004	SSI,	AUS	1	2001-2002	Other	Int (S)	2	Low	Low	Low	Low	Low	Low	Low	Low
Curtis, 2004	SSI,	AUS	1		Other	Int (S)		Low	Low	Uncl	High	Low	Low	Low	High
Daneman, 2011	SSI,	CAN	1	2008-2009	CDC NHSN	Int (S)	1	Uncl	Low	Low	Low	Low	Low	Low	Uncl
Drees, 2010	VAP,	USA	1	2007-2008	CDC NHSN	Int (S)	1	Low	Low	Low	Uncl	Low	Low	Low	Low

Table S3 | *Continued*

Author, year	HAI studied	Loc	N Center	Study period	definition	Intend appl	N	Risk of bias				Applicability			RoB
								Pat sel	Index test	Ref	Flow	Pat sel	Index test	Ref	PPV
Gerbier, 2011	SSI, BSI, CLABSI, UTI, VAP, Pneu,	FR	1	2000-2007	Other	Int (S)	1	Low	Low	Low	Uncl	Low	Low	Low	Low
Haley, 2012	SSI,	USA	176	2008-2010	CDC NHSN	Ext	2	Low	Uncl	Low	Low	Low	Low	Low	Low
Haley, 2012	SSI,	USA	176		CDC NHSN	Ext		Low	Uncl	High	High	Low	Low	Low	High
Hebden, 2000	SSI,	USA	1	1997	CDC NNIS	Int (S)	1	Low	Low	Low	Low	Low	Low	Low	Low
Heisler, 2009	UTI, CAUTI,	USA	1	2004-2005	Clinical	Int (S)	1	Low	Low	High	Uncl	Low	Low	Uncl	High
Hollenbeak, 2011	SSI,	USA	1	2007-2008	CDC NHSN	Int (S)	1	Low	Low	Low	Low	Low	Low	Low	Low
Hougland, 2008	BSI, Pneu	USA	77	2001-2003	Unclear	Ext	1	Low	Low	Low	Uncl	Low	Low	Uncl	Low
Huang, 2011	SSI,	USA	671	2005	CDC NHSN	Ext	3	Low	High	High	High	Low	Low	Low	High
Huang, 2011	SSI,	USA	671		Unclear	Ext		Low	Low	High	Uncl	Low	Low	High	High
Huang, 2011	SSI,	USA	671		CDC NHSN	Ext		Low	Low	High	High	Low	Low	Low	High
Inacio, 2011	SSI,	USA	?	2006-2008	CDC NHSN	Int (S)	1	Low	Low	Low	Low	Low	Low	Low	Low
Julian, 2006	SSI, VAP, UTI, CAUTI,	USA	1	2004	CDC NHSN	Ext PHC4	1	Low	Low	High	High	Low	Low	Low	High
Kanerva, 2009	SSI, BSI, UTI, VAP, Pneu,	FI	20	2005	Other	Int (S)	1	Low	Uncl	Low	Low	Low	Low	Low	Uncl
Koch, 2012	Sepsis,	USA	1	2009-2010	(VA/N)SQIP	Ext PSI 4.2	2	Low	Low	Low	Low	Low	Low	Low	Low
Koch, 2012	Sepsis,	USA	1		Other	Ext PSI 4.2		Low	Low	Low	Low	Low	Low	Low	Low
Landers, 2010	UTI,	USA	1	2007	Other	Int (S)	1	Low	Low	Low	Low	Low	Low	High	Low
Lawson, 2012	SSI, Sepsis, Pneu, UTI,	USA	214	2005-2008	(VA/N)SQIP	Ext	1	Low	Uncl	Low	Low	Low	Low	Low	Uncl

Table S3 | *Continued*

Author, year	HAI studied	Loc	N Center	Study period	definition	Intend appl	N	Risk of bias Pat sel	Index test	Ref	Flow	Applicability Pat sel	Index test	Ref	RoB PPV
Lee, 2011	SSI, BSI, Pneu, UTI,	JP	4	2005-2009	CDC NHSN	Int (C) PHC4	1	Low	Low	Low	Low	High	High	Low	Low
Leth, 2006	SSI,	DK	1	1999-2002	CDC NHSN	Int (C)	2	Low	Uncl	Low	Low	Low	High	Low	Low
Leth, 2006	SSI,	DK	1	1999-2002	CDC NHSN	Int (C)		Uncl	Low	Uncl	Low	Low	Low	Low	Uncl
Leth, 2010	SSI	DK	3	2007-2008	CDC NHSN	Int (C)	1	Low	Low	Low	High	Low	High	Low	High
Meddings, 2010	CAUTI,	USA	1	2006-2007	Other	Ext HAC	1	Low	Low	High	High	Low	Low	High	High
Miner, 2004	SSI,	USA	7	1996-1999	CDC NNIS	Int (C)	1	Low	Low	High	High	Low	High	Low	High
Moehring, 2013	CLABSI,	USA	3	2007-2009	CDC NHSN	Ext HAC	1	Low	Low	Low	High	Low	Low	Low	Low
Moro, 2004	SSI,	IT	31	2001	CDC NNIS	Int (S)	1	Low	Uncl	Low	Low	Low	Low	Low	Uncl
Murff, 2011	Sepsis, Pneu	USA	6	1999-2006	(VA/N)SQIP	Ext PSI 3.1	1	Low	Low	Low	Low	Low	Low	Low	Low
Ollendorf, 2002	Sepsis,	USA	10	Uncl	Clinical	Int (S)	1	Uncl	Uncl	Low	Low	Uncl	Low	High	Uncl
Olsen, 2010	SSI,	USA	1	1998-2002	CDC NHSN	Int (S)	1	Uncl	Low	High	High	Low	Low	Low	High
Platt, 2002	SSI,	USA	4	1996-1999	CDC NNIS	Int (C)	1	Uncl	Low	High	High	Low	High	Low	High
Pokorny, 2006	CLABSI, VAP, CAUTI,	ESP	1	1999-2002	CDC NHSN	Int (C)	1	Low	Uncl	Low	Low	Low	High	Uncl	Uncl
Romano, 2009	Sepsis,	USA	110	2000-2001	(VA/N)SQIP	Ext PSI 2.1	2	Low	Low	Low	Low	Low	Low	Low	Low
Romano, 2009	Sepsis,	USA	110	2000-2001	(VA/N)SQIP	Ext PSI 2.1		Low	High	Low	Low	Low	Low	Low	High
Sands, 2003	SSI,	USA	5	1995-1997	CDC NNIS	Int (C)	1	Uncl	Low	High	High	Low	High	Low	High
Scanlon, 2008	CLABSI, Sepsis,	USA	28	2003-2005	Other	Ext PDI	1	Low	Low	High	PPV	Low	Low	High	High

Table S3 | *Continued*

Author, year	HAI studied	Loc	N Center	Study period	definition	Intend appl	N	Risk of bias Pat sel	Index test	Ref	Flow	Applicability Pat sel	Index test	Ref	RoB PPV
Sherman, 2006	SSI, CLABSI, VAP, CAUTI,	USA	1	2004	CDC NHSN	Ext PHC4	1	Low	Low	High	High	Low	Low	Low	High
Song, 2008	SSI,	USA	1	2005	CDC NNIS	Int (C)	1	Low	Uncl	Low	Low	Low	High	Low	Uncl
Spolaore, 2005	SSI,	IT	3	2001	CDC NHSN	Int (C)	1	Low	Low	High	PPV	Low	High	Low	High
Stamm, 2012	CLABSI, VAP, CAUTI,	USA	1	2009	CDC NHSN	Int (S)	1	Low	Uncl	Uncl	High	Low	Low	Low	High
Stevenson, 2008	SSI, CLABSI, VAP,	USA	1	2005	CDC NHSN	Ext PHC4	2	Low	Low	Low	Low	Low	Low	Low	Low
Stevenson, 2008	SSI, CLABSI, VAP,	USA	1	2005	CDC NHSN	Ext PHC4		Low	Low	Uncl	High	Low	Low	Low	High
Stone, 2007	CLABSI,	USA	24	2002	CDC NHSN	Ext PSI 2.1	1	Low	Low	Low	Low	Low	Low	Low	Low
Tehrani, 2013	CLABSI,	USA	6	2009-2011	CDC NHSN	Ext HAC	2	Low	Low	Low	Low	Low	Low	Low	Low
Tehrani, 2013	CLABSI,	USA	6	2009 -2011	CDC NHSN	Ext HAC		Low	Low	Uncl	PPV	Low	Low	Low	Low
Tinelli, 2011	SSI, UTI,	USA	28	2005-2006	CDC NHSN	Int (S)	1	Low	Uncl	Low	Low	Low	Low	Low	Uncl
van Mourik, 2013	Drain-related meningitis	NL	1	2004-2010	CDC NHSN	Int (S)	1	Uncl	Low	Low	Low	Low	Low	Low	Uncl
Verelst, 2010	SSI, Sepsis, VAP,	BE	8	2005	Clinical	Ext PSI 3.1	1	High	Low	Low	Uncl	Low	Low	Low	High
Yokoe, 2001	Postpartum	USA	1	1993-1995	CDC NNIS	Int (C)	1	Low	Low	High	High	Low	High	Low	High
Yokoe, 2004	SSI,	USA	13	1998-2001	CDC NNIS	Int (C)	2	Low	Low	High	High	Low	High	Low	High
Yokoe, 2004	SSI,	USA	13	1998-2001	CDC NNIS	Int (C)		Low	Low	High	Uncl	Low	High	Low	High
Yokoe, 2012	SSI,	USA	5	2003-2005	CDC NHSN	Int (C)	1	Low	Low	Uncl	Low	Low	High	Low	Low

Table S3 | *Continued*

Author, year	HAI studied	Loc	N Center	Study period	definition	Intend appl	N	Risk of bias Pat sel	Index test	Ref	Flow	Applicability Pat sel	Index test	Ref	RoB PPV
Zhan, 2009	CAUTI,	USA	uncl	2005-2006	Other	Ext	1	Uncl	Uncl	Low	Low	Low	Low	Uncl	Uncl
Zrelak, 2011	CLABSI,	USA	23	2005	CDC NHSN	Ext PSI 3.1	1	Low	Low	High	PPV	Low	Low	Low	High

Abbreviations & Legend

HAI types: (CA)UTI – (catheter-associated) urinary tract infection, (CLA)BSI – central-line associated bloodstream infection, Pneu – pneumonia, SSI – surgical site infection, VAP – ventilator-associated pneumonia.

Loc (Country): AUS – Australia, BE – Belgium, CAN – Canada, DK – Denmark, ESP – Spain, FI – Finland, FR – France, IT-Italy, JP – Japan, NL – Netherlands, USA – United States of America,

Definition: CDC-NHSN or CDC-NNIS – definitions from the Centers for Disease Control Healthcare Safety Network or its predecessor, (VA/N)SQIP – definitions & methods from the National (or Veteran's Affairs) Surgical Quality Improvement Project.

Intend appl: Intended application of administrative data within HAI surveillance.

Ext – for external quality assessment, e.g. public reporting or pay-for-performance.

Int (S) – to support within hospital surveillance as sole method of finding possible HAI cases.

Int (C) – to support within hospital surveillance, combined with other indicators of HAI.

If applicable, specific metrics are indicated: HAC – Healthcare-associated condition as defined by the Centers for Medicare and Medicaid Services, PHC4 – code selection specified by the Pennsylvania Healthcare Cost Containment Council, PSI – Patient Safety Indicator.

N: design number (if more than one study design included within each study)

Risk of bias (Rob) & applicability: Patient selection (Pat Sel), Index test, Reference standard (Ref) and Flow. If a study assesses only the positive predictive value (partial verification, fully dependent on the index test – e.g. administrative data), and the risk of bias of the on the flow domain is low for the PPV estimate, these studies have been marked as "PPV" in the risk of bias on flow column. The overall risk of bias of the PPV estimate is marked in RoB PPV column.

Notes:

The following studies used the ICD-10 coding system: Curtis 2004, Daneman 2011, Gerbier 2011, Kanerva 2009, Lee 2011, Leth 2006, Leth 2010. Heisler 2009 used a different coding system.

In the following studies a present-on-admission indicator was explicitly included in the administrative data algorithm:
Cima 2011, Haley 2012, Koch 2012, Meddings 2010, Moehring 2013, Murff 2011, Tehrani 2013, Zrelak 2011

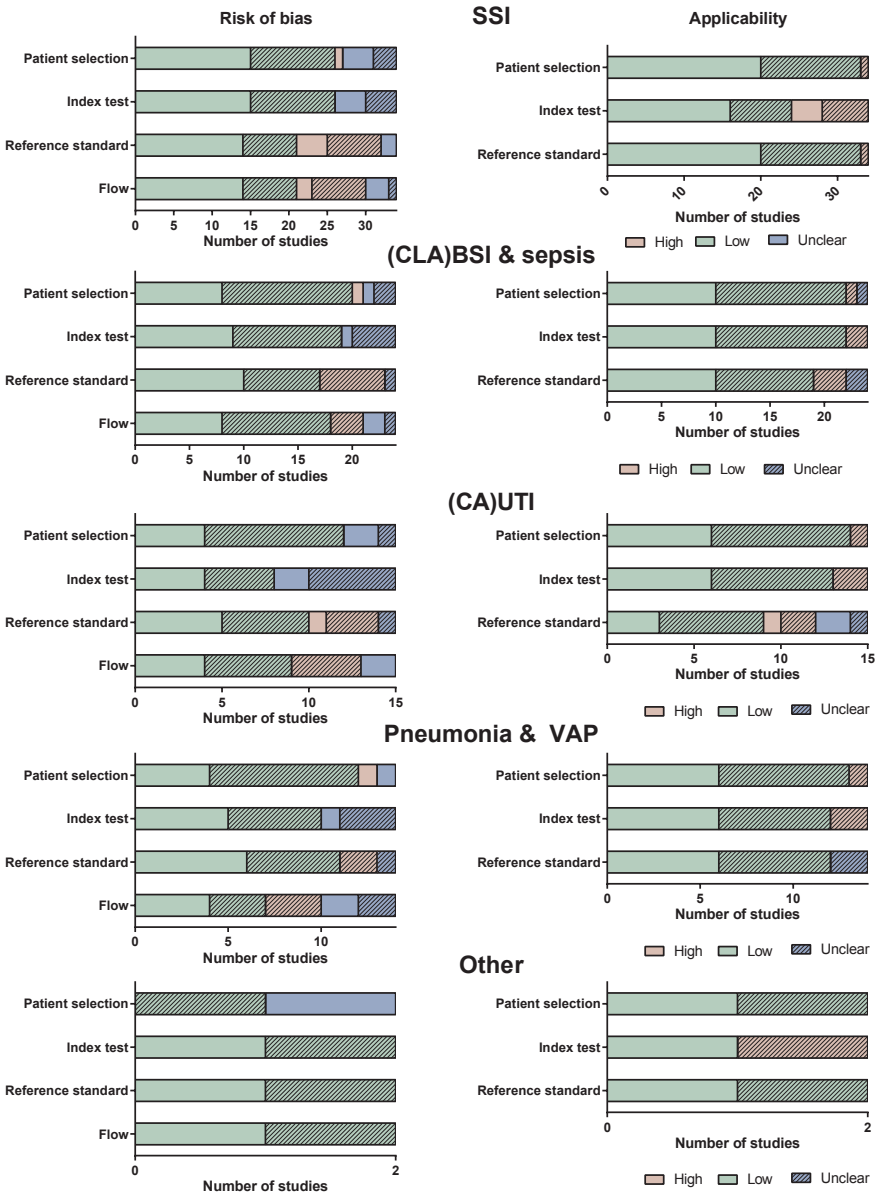


Figure S1 | Summary risk of bias, by HAI type. Summary of risk of bias and applicability for all studies (n = 57), assessed using the *Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2)* methods.

Some studies contain multiple comparisons; in this case the lowest risk of bias per study is included. Shading denotes studies where extraction of complete two-by-two tables was not possible, including studies only assessing positive predictive values.

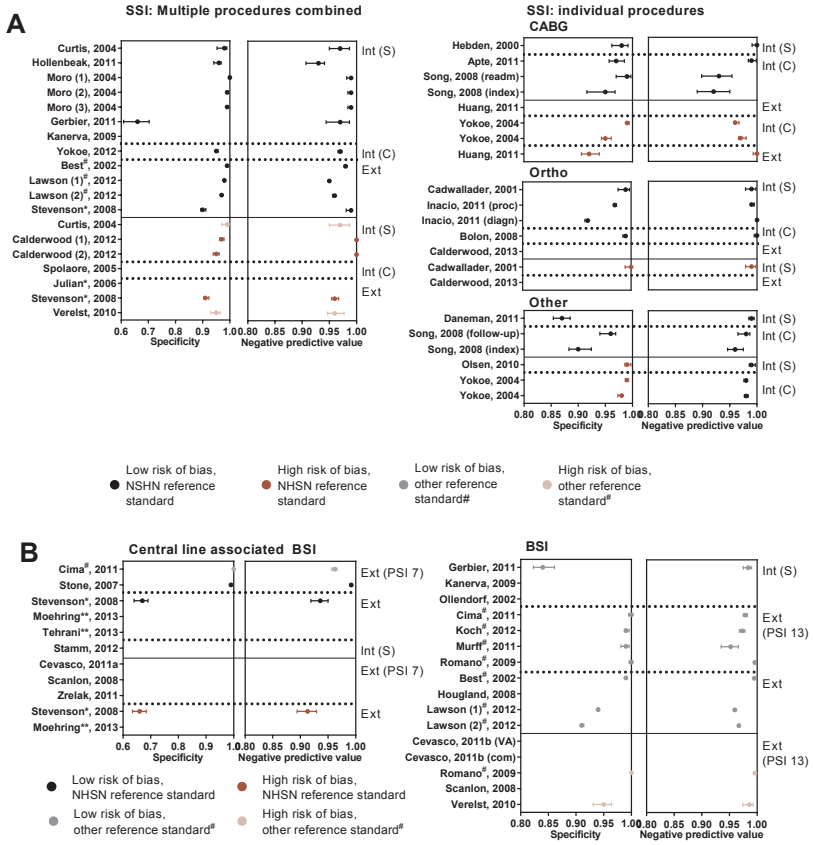
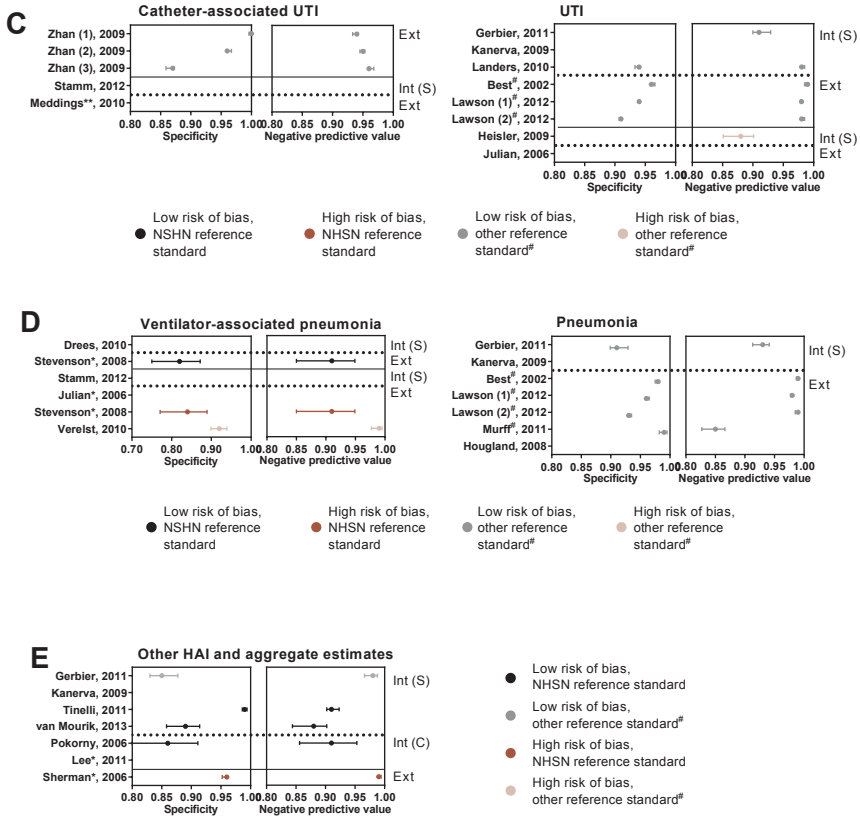


Figure S2 | Forest plots for specificity and negative predictive value, stratified by HAI type and relevant study characteristics.

A. Surgical site infection, B. (Catheter-associated) bloodstream infection, C. (Catheter-associated) urinary tract infection, D. (Ventilator-associated) pneumonia. E. Other HAI or studies presenting only data aggregated for multiple types of infection.



Studies are grouped by the intended application of administrative data:
Int (S) – used in isolation to support within-hospital surveillance efforts.
Int (C) – used to support within-hospital surveillance, combined with other indicators of infection.
Ext – used for external quality assessment, including public reporting and pay-for-performance.
In studies including multiple specifications of the administrative data algorithm, these are numbered sequentially. 95% confidence intervals are derived using the exact binomial method.
Notes: In studies including multiple specifications of the administrative data algorithm, these are numbered sequentially. 95% confidence intervals are derived using the exact binomial method. If multiple study designs were performed within a single study, they are mentioned separately.
#: reference standard from Surgical Quality Improvement Project (NSQIP or VASQIP). *: Code selection based on specification from Pennsylvania Health Cost Containment Council. **: HAC specification.
Abbreviations: BSI – bloodstream infection, CABG – coronary artery bypass graft, DRM – drain-related meningitis, Ortho – orthopedic procedure, PSI – Patient safety indicator, Sep – Sepsis, SSI – surgical site infection, UTI – urinary tract infection.

REFERENCE LIST

1. Panzer RJ, Gitomer RS, Greene WH, Webster PR, Landry KR, Riccobono CA. Increasing demands for quality measurement. *JAMA.* **2013**;310:1971-1980.
2. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* **1998**;36:8-27.
3. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care.* **2003**;12 Suppl 2:ii58-ii63.
4. Jarman B, Pieter D, van der Veen AA et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Health Care.* **2010**;19:9-13.
5. Glance LG, Osler TM, Mukamel DB, Meredith W, Wagner J, Dick AW. TMPM-ICD9: a trauma mortality prediction model based on ICD-9-CM codes. *Ann Surg.* **2009**;249:1032-1039.
6. Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med.* **2007**;357:1573-1575.
7. Agency for Healthcare Research and Quality. Patient Safety Indicators Overview. **2014**. http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx Last accessed: Mar 18 2014.
8. Department of Health and Human Services, Centers for Medicare and Medicaid Services. FY 2014 IPPS Final Rule Medicare Program: Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status. *Federal Register.* **2013**;78:50495-51040.
9. Centers for Medicare and Medicaid Services. Press release: CMS issues proposed inpatient hospital payment system regulation. **2014**. http://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-releases/2014-Press-releases-items/2014-04-30.html Last accessed: May 27 2014.
10. Drosler SE, Romano PS, Tancredi DJ, Klazinga NS. International comparability of patient safety indicators in 15 OECD member countries: a methodological approach of adjustment by secondary diagnoses. *Health Serv Res.* **2012**;47:275-292.
11. Woeltje KF. Moving into the future: electronic surveillance for healthcare-associated infections. *J Hosp Infect.* **2013**;84:103-105.
12. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of Administrative Code Data for the Surveillance of Healthcare-Associated Infections: A Systematic Review and Meta-Analysis. *Clin Infect Dis.* **2014**;58:688-696.
13. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. *Clin Infect Dis.* **2009**;49:949-955.
14. Farmer SA, Black B, Bonow RO. Tension between quality measurement, public quality reporting, and pay for performance. *JAMA.* **2013**;309:349-350.
15. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med.* **1997**;127:666-674.
16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **2009**;6:e1000097.
17. Whiting PF, Rutjes AW, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* **2011**;155:529-536.
18. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol.* **2011**;64:821-829.
19. Naaktgeboren CA, de Groot JA, van SM, Moons KG, Reitsma JB. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med.* **2013**;159:195-202.
20. Macaskill P, Gatsonis C, Deeks JJ, Bossuyt PM, Harbord R, Takwoingi Y. Chapter 10: Analysing and Presenting the Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration; **2010**.
21. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* **2005**;58:982-990.

22. Apte M, Landers T, Furuya Y, Hyman S, Larson E. Comparison of Two Computer Algorithms to Identify Surgical Site Infections. *Surg Infect (Larchmt).* **2011**;12:459-464.
23. Best WR, Khuri SF, Phelan M et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg.* **2002**;194:257-266.
24. Bolon MK, Hooper D, Stevenson KB et al. Improved surveillance for surgical site infections after orthopedic implantation procedures: extending applications for automated data. *Clin Infect Dis.* **2009**;48:1223-1229.
25. Braun BI, Kritchevsky SB, Kusek L et al. Comparing bloodstream infection rates: the effect of indicator specifications in the evaluation of processes and indicators in infection control (EPIC) study. *Infect Control Hosp Epidemiol.* **2006**;27:14-22.
26. Cadwallader HL, Toohey M, Linton S, Dyson A, Riley TV. A comparison of two methods for identifying surgical site infections following orthopaedic surgery. *J Hosp Infect.* **2001**;48:261-266.
27. Calderwood MS, Ma A, Khan YM et al. Use of Medicare Diagnosis and Procedure Codes to Improve Detection of Surgical Site Infections following Hip Arthroplasty, Knee Arthroplasty, and Vascular Surgery. *Infect Control Hosp Epidemiol.* **2012**;33:40-49.
28. Calderwood MS, Kleinman K, Bratzler DW et al. Use of Medicare Claims to Identify US Hospitals with a High Rate of Surgical Site Infection after Hip Arthroplasty. *Infect Control Hosp Epidemiol.* **2013**;34:31-39.
29. Campbell PG, Malone J, Yadla S et al. Comparison of ICD-9-based, retrospective, and prospective assessments of perioperative complications: assessment of accuracy in reporting. *J Neurosurg Spine.* **2011**;14:16-22.
30. Cevasco M, Borzecki AM, O'Brien WJ et al. Validity of the AHRQ Patient Safety Indicator "central venous catheter-related bloodstream infections". *J Am Coll Surg.* **2011**;212:984-990.
31. Cevasco M, Borzecki AM, Chen Q et al. Positive predictive value of the AHRQ Patient Safety Indicator "Postoperative Sepsis": implications for practice and policy. *J Am Coll Surg.* **2011**;212:954-961.
32. Cima RR, Lackore KA, Nehring SA et al. How best to measure surgical quality? Comparison of the Agency for Healthcare Research and Quality Patient Safety Indicators (AHRQ-PSI) and the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) postoperative adverse events at a single institution. *Surgery.* **2011**;150:943-949.
33. Curtis M, Graves N, Birrell F et al. A comparison of competing methods for the detection of surgical-site infections in patients undergoing total arthroplasty of the knee, partial and total arthroplasty of hip and femoral or similar vascular bypass. *J Hosp Infect.* **2004**;57:189-193.
34. Daneman N, Ma X, Eng-Chong M, Callery S, Guttmann A. Validation of administrative population-based data sets for the detection of cesarean delivery surgical site infection. *Infect Control Hosp Epidemiol.* **2011**;32:1213-1215.
35. Drees M, Hausman S, Rogers A, Freeman L, Frosch K, Wroten K. Underestimating the impact of ventilator-associated pneumonia by use of surveillance data. *Infect Control Hosp Epidemiol.* **2010**;31:650-652.
36. Gerbier S, Bouzbid S, Pradat E et al. [Use of the French medico-administrative database (PMSI) to detect nosocomial infections in the University hospital of Lyon.]. *Rev Epidemiol Sante Publique.* **2011**;59:3-14.
37. Haley VB, Van AC, Tserenpuntsag B et al. Use of administrative data in efficient auditing of hospital-acquired surgical site infections, new york state 2009-2010. *Infect Control Hosp Epidemiol.* **2012**;33:565-571.
38. Hebden J. Use of ICD-9-CM coding as a case-finding method for sternal wound infections after CABG procedures. *Am J Infect Control.* **2000**;28:202-203.
39. Heisler CA, Melton LJ, III, Weaver AL, Gebhart JB. Determining perioperative complications associated with vaginal hysterectomy: code classification versus chart review. *J Am Coll Surg.* **2009**;209:119-122.
40. Hollenbeak CS, Boltz MM, Nikkel LE, Schaefer E, Ortenzi G, Dillon PW. Electronic measures of surgical site infection: implications for estimating risks and costs. *Infect Control Hosp Epidemiol.* **2011**;32:784-790.
41. Hougland P, Nebeker J, Pickard S et al. Using ICD-9-CM Codes in Hospital Claims Data to Detect Adverse Events in Patient Safety Surveillance. In: Hendrikson K, Battles JB, Keyes A, Grady ML, eds. *Advances in Patient Safety: New Directions and Alternative Approaches*. Rockville (MD): Agency for Healthcare Research and Quality; **2008**.
42. Huang SS, Placzek H, Livingston J et al. Use of Medicare Claims to Rank Hospitals by Surgical Site Infection Risk following Coronary Artery Bypass Graft Surgery. *Infect Control Hosp Epidemiol.* **2011**;32:775-783.

43. Inacio MCS, Paxton EW, Chen Y et al. Leveraging electronic medical records for surveillance of surgical site infection in a total joint replacement population. *Infect Control Hosp Epidemiol.* **2011**;32:351-359.
44. Julian KG, Brumbach AM, Chicora MK et al. First year of mandatory reporting of healthcare-associated infections, Pennsylvania: an infection control-chart abstractor collaboration. *Infect Control Hosp Epidemiol.* **2006**;27:926-930.
45. Kanerva M, Ollgren J, Virtanen MJ, Lyytikainen O. Estimating the annual burden of health care-associated infections in Finnish adult acute care hospitals. *Am J Infect Control.* **2009**;37:227-230.
46. Koch CG, Li L, Hixson E, Tang A, Phillips S, Henderson JM. What are the real rates of postoperative complications: elucidating inconsistencies between administrative and clinical data sources. *J Am Coll Surg.* **2012**;214:798-805.
47. Landers T, Apte M, Hyman S, Furuya Y, Glied S, Larson E. A comparison of methods to detect urinary tract infections using electronic data. *Jt Comm J Qual Patient Saf.* **2010**;36:411-417.
48. Lawson EH, Louie R, Zingmond DS et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg.* **2012**;256:973-981.
49. Lee J, Imanaka Y, Sekimoto M, Nishikawa H, Ikai H, Motohashi T. Validation of a novel method to identify healthcare-associated infections. *J Hosp Infect.* **2011**;77:316-320.
50. Leth RA, Moller JK. Surveillance of hospital-acquired infections based on electronic hospital registries. *J Hosp Infect.* **2006**;62:71-79.
51. Leth RA, Norgaard M, Uldbjerg N, Thomsen RW, Moller JK. Surveillance of selected post-caesarean infections based on electronic registries: validation study including post-discharge infections. *J Hosp Infect.* **2010**;75:200-204.
52. Meddings J, Saint S, McMahon LF, Jr. Hospital-acquired catheter-associated urinary tract infection: documentation and coding issues may reduce financial impact of Medicare's new payment policy. *Infect Control Hosp Epidemiol.* **2010**;31:627-633.
53. Miner AL, Sands KE, Yokoe DS et al. Enhanced identification of postoperative infections among outpatients. *Emerg Infect Dis.* **2004**;10:1931-1937.
54. Moehring RW, Staheli R, Miller BA, Chen LF, Sexton DJ, Anderson DJ. Central Line-Associated Infections as Defined by the Centers for Medicare and Medicaid Services' Hospital-Acquired Condition versus Standard Infection Control Surveillance: Why Hospital Compare Seems Conflicted. *Infect Control Hosp Epidemiol.* **2013**;34:238-244.
55. Moro ML, Morsillo F. Can hospital discharge diagnoses be used for surveillance of surgical-site infections? *J Hosp Infect.* **2004**;56:239-241.
56. Murff HJ, FitzHenry F, Matheny ME et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA.* **2011**;306:848-855.
57. Ollendorf DA, Fendrick AM, Massey K, Williams GR, Oster G. Is sepsis accurately coded on hospital bills? *Value Health.* **2002**;5:79-81.
58. Olsen MA, Fraser VJ. Use of diagnosis codes and/or wound culture results for surveillance of surgical site infection after mastectomy and breast reconstruction. *Infect Control Hosp Epidemiol.* **2010**;31:544-547.
59. Platt R, Kleinman K, Thompson K et al. Using automated health plan data to assess infection risk from coronary artery bypass surgery. *Emerg Infect Dis.* **2002**;8:1433-1441.
60. Pokorny L, Rovira A, Martin-Baranera M, Gimeno C, onso-Tarres C, Vilarasau J. Automatic detection of patients with nosocomial infection by a computer-based surveillance system: a validation study in a general hospital. *Infect Control Hosp Epidemiol.* **2006**;27:500-503.
61. Romano PS, Mull HJ, Rivard PE et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. *Health Serv Res.* **2009**;44:182-204.
62. Sands KE, Yokoe DS, Hooper DC et al. Detection of postoperative surgical-site infections: comparison of health plan-based surveillance with hospital-based programs. *Infect Control Hosp Epidemiol.* **2003**;24:741-743.
63. Scanlon MC, Harris JM, Levy F, Sedman A. Evaluation of the agency for healthcare research and quality pediatric quality indicators. *Pediatrics.* **2008**;121:e1723-e1731.
64. Sherman ER, Heydon KH, St John KH et al. Administrative data fail to accurately identify cases of healthcare-associated infection. *Infect Control Hosp Epidemiol.* **2006**;27:332-337.

65. Song X, Cosgrove SE, Pass MA, Perl TM. Using hospital claim data to monitor surgical site infections for inpatient procedures. *Am J Infect Control.* **2008**;36:S32-S36.
66. Spolaore P, Pellizzer G, Fedeli U et al. Linkage of microbiology reports and hospital discharge diagnoses for surveillance of surgical site infections. *J Hosp Infect.* **2005**;60:317-320.
67. Stamm AM, Bettacchi CJ. A comparison of 3 metrics to identify health care-associated infections. *Am J Infect Control.* **2012**;40:688-691.
68. Stevenson KB, Khan Y, Dickman J et al. Administrative coding data, compared with CDC/NHSN criteria, are poor indicators of health care-associated infections. *Am J Infect Control.* **2008**;36:155-164.
69. Stone PW, Horan TC, Shih HC, Mooney-Kane C, Larson E. Comparisons of health care-associated infections identification using two mechanisms for public reporting. *Am J Infect Control.* **2007**;35:145-149.
70. Tehrani DM, Russell D, Brown J et al. Discord among Performance Measures for Central Line-Associated Bloodstream Infection. *Infect Control Hosp Epidemiol.* **2013**;34:176-183.
71. Tinelli M, Mannino S, Lucchi S et al. Healthcare-acquired infections in rehabilitation units of the Lombardy Region, Italy. *Infection.* **2011**;39:353-358.
72. van Mourik MS, Troelstra A, Moons KG, Bonten MJ. Accuracy of hospital discharge coding data for the surveillance of drain-related meningitis. *Infect Control Hosp Epidemiol.* **2013**;34:433-436.
73. Verelst S, Jacques J, Van den HK et al. Validation of Hospital Administrative Dataset for adverse event screening. *Qual Saf Health Care.* **2010**;19:e25.
74. Yokoe DS, Christiansen CL, Johnson R et al. Epidemiology of and surveillance for postpartum infections. *Emerg Infect Dis.* **2001**;7:837-841.
75. Yokoe DS, Noskin GA, Cunnigham SM et al. Enhanced identification of postoperative infections among inpatients. *Emerg Infect Dis.* **2004**;10:1924-1930.
76. Yokoe DS, Khan Y, Olsen MA et al. Enhanced surgical site infection surveillance following hysterectomy, vascular, and colorectal surgery. *Infect Control Hosp Epidemiol.* **2012**;33:768-773.
77. Zhan C, Elixhauser A, Richards CL, Jr. et al. Identification of hospital-acquired catheter-associated urinary tract infections from Medicare claims: sensitivity and positive predictive value. *Med Care.* **2009**;47:364-369.
78. Zrelak PA, Sadeghi B, Utter GH et al. Positive predictive value of the Agency for Healthcare Research and Quality Patient Safety Indicator for central line-related bloodstream infection ("selected infections due to medical care"). *J Healthc Qual.* **2011**;33:29-36.
79. Meddings JA, Reichert H, Rogers MA, Saint S, Stephansky J, McMahon LF. Effect of nonpayment for hospital-acquired, catheter-associated urinary tract infection: a statewide analysis. *Ann Intern Med.* **2012**;157:305-312.
80. Zhan C, Elixhauser A, Richards CL, Jr. et al. Identification of hospital-acquired catheter-associated urinary tract infections from Medicare claims: sensitivity and positive predictive value. *Med Care.* **2009**;47:364-369.
81. Safdar N, Anderson DJ, Braun BI et al. The evolving landscape of healthcare-associated infections: recent advances in prevention and a road map for research. *Infect Control Hosp Epidemiol.* **2014**;35:480-493.
82. Patrick SW, Davis MM, Sedman AB et al. Accuracy of hospital administrative data in reporting central line-associated bloodstream infections in newborns. *Pediatrics.* **2013**;131 Suppl 1:S75-S80.
83. Letourneau AR, Calderwood MS, Huang SS, Bratzler DW, Ma A, Yokoe DS. Harnessing Claims to Improve Detection of Surgical Site Infections following Hysterectomy and Colorectal Surgery. *Infect Control Hosp Epidemiol.* **2013**;34:1321-1323.
84. Knepper BC, Young H, Jenkins TC, Price CS. Time-saving impact of an algorithm to identify potential surgical site infections. *Infect Control Hosp Epidemiol.* **2013**;34:1094-1098.
85. Quan H, Eastwood C, Cunningham CT et al. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ Open.* **2013**;3:e003716.
86. Atchley KD, Pappas JM, Kennedy AT et al. Use of Administrative Data for Surgical Site Infection Surveillance After Congenital Cardiac Surgery Results in Inaccurate Reporting of Surgical Site Infection Rates. *Ann Thorac Surg.* **2014**;97:651-657.
87. Cass AL, Kelly JW, Probst JC, Addy CL, McKeown RE. Identification of device-associated infections utilizing administrative data. *Am J Infect Control.* **2013**;41:1195-1199.

88. Leclere B, Lasserre C, Bourigault C et al. Computer-enhanced surveillance of surgical site infections: early assessment of a generaliable method for French hospitals. *European Conference of Clinical Microbiology and Infectious Diseases*. **2014**.

89. Mull HJ, Borzecki AM, Loveland S et al. Detecting adverse events in surgery: comparing events detected by the Veterans Health Administration Surgical Quality Improvement Program and the Patient Safety Indicators. *Am J Surg*. **2014**;207:584-595.

90. Thompson ND, Yeh LL, Magill SS, Ostroff SM, Fridkin SK. Investigating Systematic Misclassification of Central Line-Associated Bloodstream Infection (CLABSI) to Secondary Bloodstream Infection During Health Care-Associated Infection Reporting. *Am J Med Qual*. **2013**;28:56-59.

91. Lindenauer PK, Lagu T, Shieh MS, Pekow PS, Rothberg MB. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. *JAMA*. **2012**;307:1405-1413.

92. Calderwood MS, Kleinman K, Soumerai SB et al. Impact of Medicare's Payment Policy on Mediastinitis Following Coronary Artery Bypass Graft Surgery in US Hospitals. *Infect Control Hosp Epidemiol*. **2014**;35:144-151.

93. Cram P, Bozic KJ, Lu X, Li Y. Use of Present-On-Admission Indicators for Complications After Total Knee Arthroplasty: An Analysis of Medicare Administrative Data. *J Arthroplasty*. **2014**;29:923-928.

94. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD -10) Version for 2010. **2010**

95. Drosler SE, Romano PS, Sundararajan V et al. How many diagnosis fields are needed to capture safety events in administrative data? Findings and recommendations from the WHO ICD-11 Topic Advisory Group on Quality and Safety. *Int J Qual Health Care*. **2014**;26:16-25.

96. de Groot JA, Bossuyt PM, Reitsma JB et al. Verification problems in diagnostic accuracy studies: consequences and solutions. *BMJ*. **2011**;343:d4770.

Part 3

Synthesis

10|

General discussion: Positioning automated surveillance of healthcare-associated infections using routine care data

Maaike S.M. van Mourik

INTRODUCTION

Automated surveillance of healthcare-associated infections (HAI) using routine care data has been advocated as a means to improve current surveillance practices and has been associated with successful infection prevention programs[1-4]. Traditional HAI surveillance - manual chart review of patient charts using standardized definitions - is hampered by limited capacity, variability in data collection methods and subjective interpretation of case-definitions, thus affecting the reliability and comparability of HAI rates across observers, time and hospitals[4-7]. Automated surveillance systems identify (possible) cases of HAI using data collected during the process of routine care stored in electronic clinical data warehouses, e.g., microbiology culture results, antibiotic dispensing or procedure codes. Implementation of these systems is expected to improve efficiency and accuracy of detection by ensuring consistent application of case-definitions.

Within the process of HAI surveillance, automated systems can have two intended applications. They may either serve as a means to support within-hospital surveillance by identifying high-risk patients requiring manual review of their medical records to confirm HAI occurrence, or be used as a quality metric directly assessing the presence or absence of HAI[8]. Whereas in the former application sensitivity and perhaps efficiency are important performance characteristics, specificity and a low rate of false-positives may be more highly valued in the latter. Reliable surveillance of HAI is a vital component of quality improvement programs by identifying opportunities for improvement and measuring the effect of interventions. In addition, as HAI rates are increasingly considered as quality indicators for benchmarking, public reporting or reimbursement schemes, generating rates that are robust and allow for valid comparisons across hospitals is becoming paramount.

APPROACHES TO AUTOMATED HAI SURVEILLANCE USING ROUTINE CARE DATA

Several possible approaches to automated HAI surveillance have been developed that differ in the type of routine care data they employ, the methods they use to detect a case of HAI and their performance characteristics[9,10].

Types of routine care data

Routine care data can be described on several dimensions. It may either be clinical data measured directly from the patient, or it can be derived from classification systems developed for other purposes, such as administrative data containing discharge codes.

Alternatively, routine care data may be classified as descriptive data representing a patient's clinical state based on diagnostic testing results or physicians' narratives versus medico-administrative data on treatments and interventions that the patient received (e.g., medication use, procedure codes)[10].

The validity of using administrative data for the purpose of HAI surveillance, and quality evaluation in general, has been questioned by many[11-14]. Their limited value was also confirmed by the studies in this thesis, both for drain-related meningitis as well as for other types of HAI (Ch. 8 & 9). However, as administrative data are generally easy to collect and fairly standardized across hospitals, albeit for different reasons than HAI surveillance, their use is still widespread, in particular in direct quality metrics[15,16]. In addition, the use of administrative data to support within-hospital surveillance efforts by enhancing case-finding is continuing to be advocated, either alone or in combination with other types of (clinical) routine care data[1,17-19]. Given the limited and inconsistent accuracy of discharge diagnoses for HAI detection, robust and ongoing validation studies are imperative if the use of administrative data is to be continued for the purpose of quality assessment (Ch. 9).

As an alternative to administrative data, clinical data - either descriptive or reflecting treatment(s) - provides valuable information necessary to identify the development of a HAI. In the surveillance approach targeting drain-related meningitis, combining diagnostic testing results with treatment information could accurately identify possible cases of meningitis (Ch. 2-5), as was also seen for other types of HAI[20-22]. Some algorithms combine clinical and administrative data, for example antibiotic use or microbiology culture results with discharge codes, to enhance sensitivity of detection; however, the incremental value of adding administrative data is generally modest[23-25]. Most systems that employ clinical data as a (main) source of information to detect HAI were primarily developed to support within-hospital surveillance efforts although clinical information is gradually becoming more prominent in outcome measures designed as quality metrics[1,26,27]. The scope of automated surveillance using clinical data may thus expand to multicenter implementation and generation of HAI rates that are meant to be comparable across hospitals.

The movement towards the use of clinical data in quality assessment metrics is strengthened by the increasing availability of clinical data through electronic data repositories. Policy makers are providing incentives to encourage the adoption of electronic health records and their 'meaningful use'[28], and clinical data warehouses are being developed in many hospitals to support clinical research and healthcare management[29]. The

generally increased attention for the re-use of routine care data has spurred pleas for standardization of data collection and its interpretation, and particular attention must be paid to ensuring valid and comparable interpretation over time and place. This will require considerable efforts not only with regard to the collection of the data itself but also to achieve structured documentation of contextual information surrounding the generation and extraction of data[30-33].

Methods of HAI identification using routine care data

A range of methods can be applied to routine care data to support HAI surveillance; most approaches aim to identify high-risk patients requiring manual chart review. The most straightforward method is to flag patients based on the presence of one marker of infection, e.g., a specific discharge code. Multiple sources of routine care data can also be combined in classification algorithms or incorporated in multivariable regression models. Alternatively, outside the scope of this thesis, fuzzy logic, neural networks or machine learning methods can be used to detect likely episodes of HAI[34,35]. As can be expected, and as found in the studies in this thesis, more complex models give accurate results and have a favorable balance between sensitivity and efficiency of HAI detection. Interestingly, the sensitivity of a surveillance system seems to be largely determined by the number of data sources screened for indications of HAI; the method chosen to interpret these data sources mainly affects the efficiency of detection (Ch. 4). For most types of HAI, there is not a single indicator that will detect all cases of infection and use of multiple data sources is a necessity[36]. Depending on the type of HAI, automated surveillance systems can achieve up to 80% workload reductions at acceptable to near-perfect sensitivity and reliability of automated systems has been favored over manual surveillance (Ch. 4, 5)[2,4,37]. Conversely, many systems that aim to serve as a direct quality metric often use administrative data in isolation, with modest sensitivity and positive predictive value (Ch. 9).

Given the growing possibilities of electronic data collection and advances in information technology, increasingly complex models are becoming more attractive. It can be hypothesized that development of more precise methods of automated surveillance will render manual confirmation of HAI status unnecessary, either because individual cases of HAI can be identified with great precision or because reliable group-level estimates can be generated, analogous to the drain-related meningitis studies. Whether ubiquitous adoption of these more intricate methods will be the way forward remains to be established, however. As models become more complicated, intuitive interpretation of the results may be lost and the incremental efficiency (precision) of using a more complicated approach

must be weighed against practical considerations. Moreover, it has yet to be determined if complex models are more or less robust to (small) differences in patient characteristics, implementation or practice variations than simpler approaches.

Importantly, irrespective of the selection of a simple or more intricate method to detect possible cases of HAI, the requirements for adequate collection of routine care data, detailed model specification and ongoing validation of surveillance methods remain. In addition, the benefit of completely replacing manual review of patient records by automated systems can be questioned as this review process may provide valuable insight on the possible etiology of HAI and opportunities for improvement, thereby contributing to the success of prevention programs.

LESSONS LEARNT, BARRIERS TO IMPLEMENTATION AND REMAINING CHALLENGES

Despite promising developments in automated surveillance of HAI using routine care data, there are hurdles to overcome prior to (large scale) introduction of such systems, from a practical perspective and stemming from the need to ensure the generation of reliable and meaningful rates.

Data collection and interpretation

First and foremost, adoption of automated surveillance systems requires the availability of high-quality data stored in an accessible format, both to identify the patient population at risk of HAI (the denominator, e.g., defined by specific procedures or use of invasive devices) and to detect cases of infection, the numerator. Unfortunately, complete availability of this data is rather an exception than the rule and as many of the current automated surveillance systems are home-built, they have limited transportability to other settings[9]. A defining feature of (the re-use of) routine care data is its dependence on clinical practice, and the absence of control over data collection[38]. As allured to above, care must be taken to ensure the comparability of routine care data interpretation, in particular when transporting systems to other settings or performing interhospital quality evaluation.

Undesirable variability in HAI detection by automated surveillance methods may result either from variation in clinical practice and hence data generating mechanisms, or from differences in implementation of algorithms. It may be argued that data on treatment(s) reflect provider behavior, and are therefore less comparable across settings

than descriptive elements such as diagnostic testing results, although the latter are also not immune to variation for either clinical or surveillance reasons (Ch. 5). In addition, missing observations – in particular for the results of diagnostic tests – are common in routine care data and are not "missing completely at random". Methods to handle these missing observations must be developed and tested in the context of HAI surveillance[39-41]. The study validating the ventilator-associated events paradigm (Ch.6) demonstrated the possible impact of differences in electronic implementation. Minor variations in data specifications had important consequences for the type of events detected and show that using electronic data in itself does not guarantee comparability of the rates obtained. When automated algorithms are used as a tool to support surveillance efforts within a single institution, methods may be adapted to local practice and these sources of variability may not be a relevant problem. Consequences may be more far reaching, however, when the aim is to generate comparable rates across hospitals.

Not unimportantly, as adoption of automated surveillance will be a staged process, data collected and interpreted electronically will need to be mirrored with care to data that is (still) being generated through manual methods as the results of both approaches may not be comparable[1,42,43]. Adaptation of quality measures to electronic data collection will perhaps be necessary, with careful specification of structured fields, adequate embedding of data collection in clinical workflow and thorough evaluation of the quality of the novel measures[44]. When developing surveillance systems using routine care data, it may be sensible to explicitly consider vulnerability to practice variation at an early stage (Ch. 5 & 7). Finally, the necessity to consider data generating mechanisms and provide detailed specifications for implementation is evidently not unique to automated HAI surveillance, but applies as strongly to traditional methods. Thus, many of the challenges we currently face will not be solved by shifting to automated surveillance using routine care data.

Other sources of variability

Regardless of the use of manual or automated surveillance methods, other sources of variation can also affect the HAI rates obtained and hamper their interpretation[45]. Two important components that deserve mentioning are post-discharge surveillance and measurement of the denominator.

Particularly for HAI that commonly develop after hospital discharge, such as surgical site infections, the duration and method of post-discharge surveillance (PDS) will affect the number of events detected. When striving for comparable interpretation of HAI rates, the trade-off between practicability and precision of surveillance calls for explicit

consideration of PDS specifications. Post-discharge surveillance may be restricted to the more serious, deep or organ-space, infections[46] or the primary treatment facility[47] and standardization of the duration of follow-up and methods of case-finding is desirable[48-50].

HAI rates are not only defined by their numerator, but also by the denominator. Methods of defining or measuring the population or population time at-risk affect the interpretation of HAI rates and may confound the measured success of a preventive strategy[51]. To illustrate, an effective prevention program that reduces the unnecessary use of urinary catheters in low-risk patients may paradoxically increase the rates of catheter-associated urinary tract infection expressed as infections per 1000 device days[52]. It has therefore been recommended to present HAI rates using a range of denominators (patients, patient days and device days) or as population-based measures that incorporate device utilization[53,54]. In addition, despite predefined specifications, the actual measurement of the denominator is also vulnerable to variability in interpretation of definitions and surveillance methods[7].

Generating meaningful rates

The process of HAI surveillance, irrespective of the method used, faces additional obstacles towards generating information that is not only reliable but also meaningful, both for the purpose of (within-hospital) quality improvement and public reporting[45].

The well-known limited reliability of manual chart review is one of the key drivers of current developments in automated surveillance methods. However, it is exactly this feature that makes assessment of novel approaches of HAI surveillance difficult. Discrepant cases or imperfect concordance between traditional and automated methods may either represent errors in ascertainment of the reference standard or inaccuracy of the novel method. In order to address these concerns, discrepancy analyses in which discordant cases are re-evaluated and reclassified are common (Ch. 9); this practice, although improving the likelihood of correct HAI detection, also risks the generation of biased performance estimates, especially when this second verification is applied to a non-random subset of cases[55]. To enable a balanced interpretation of the accuracy of novel surveillance approaches, robust study methodologies that apply the same reference standard to (a random sample of) all patients and present both results obtained before and after reclassification separately are desirable[56].

In addition, for the results of surveillance efforts to be meaningful to stakeholders, the events coined as HAI must be deemed of sufficient clinical relevance to drive

improvement initiatives and be accepted as a reflective of quality of care. In an effort to address some of the issues with reliability of HAI surveillance, the United States *National Healthcare Safety Network* is in the process of revising definitions to measure events in a more consistent and objective manner that is less vulnerable to (un)intentional gaming[1], especially by removing elements that require (subjective) chart review. A recent example is the introduction of the ventilator-associated events paradigm as a replacement of surveillance of ventilator-associated pneumonia, an infection that is notoriously difficult to assess[57,58]. Although this new approach facilitates electronic implementation, room for manipulation remains and the capacity of the events detected to drive quality improvement has been questioned due to the detection of a broad scope of clinical conditions that may not be liable to preventive measures[59-61](Ch. 6). This latter concern jeopardizes the commitment of (clinical) stakeholders that is necessary to achieve the desired quality improvement. Furthermore, the rapid introduction of the novel constructs and the simultaneous discontinuation of previous methodologies thwarts a thorough evaluation of the new methods and discards the possibility of linking new data to historical rates, especially as the novel definitions are revised frequently[26,62].

CONSIDERATIONS FOR THE ADOPTION OF (AUTOMATED) SURVEILLANCE SYSTEMS

Automated surveillance systems that employ electronic routine care data may increase the capacity and perhaps consistency of HAI detection, but sources of variation remain pervasive. Careful adoption of automated systems is warranted and modification of case-definitions to facilitate automation requires extensive validation. The recommendations proposed above are summarized in **table 1**. A recurring distinction that determines the relevance of possible concerns with HAI surveillance methods and the meaningfulness of the results is the intended use of the outcomes. Their use for within-hospital surveillance efforts and quality improvement requires less standardization across settings and benefits from surveillance practices that align closely with local clinical practices, whereas application to interfacility comparisons calls for thorough standardization. As outlined below, incorporating HAI rates in public reporting schemes or coupling them to financial incentives requires careful adjustment for underlying patient characteristics and may additionally endanger the reliability of surveillance.

Table 1 | Considerations when developing (automated) HAI surveillance methods.

Designing a surveillance method	
Intended application	Clearly define the intended application of the surveillance system: for within-hospital surveillance or as a direct quality indicator. Identify associated key measures of accuracy.
Practice variation	Explicitly consider vulnerability of the system to practice variation early in the process of development, and provide solutions for adaptation to new settings.
Other sources of variation	Address other possible sources of variation aside from HAI ascertainment, e.g., defining the population at risk, post-discharge surveillance.
Redefining HAI	When redesigning outcome definitions, carefully balance requirements with regards to feasibility of surveillance to clinical relevance of the construct as a quality measure.
Data sources in automated surveillance systems	
Type of routine care data	If collection of clinical routine care data is feasible and adequately standardized, this data may be preferable over administrative data for the purpose of HAI surveillance.
Data specification	Provide detailed specifications on data collection and implementation of the automated algorithm, and assess the robustness of the method to differences in implementation or clinical practice.
Ongoing validation	Perform ongoing validation of data collection and interpretation, irrespective of the type of data used (manual data collection, administrative data or clinical routine care data).
Implementing new methods or definitions	
Assessing accuracy	Use robust study designs to assess the accuracy of novel surveillance methods, paying particular attention to bias introduced by partial or differential verification patterns and the limitations of current HAI surveillance methods as a reference standard.
Comparative data	Ensure the generation of comparative data by maintaining parallel surveillance systems when introducing new or modified methods or definitions.

Abbrevations: HAI – healthcare-associated infection.

(AUTOMATED) HAI SURVEILLANCE IN THE CONTEXT OF QUALITY EVALUATION, PUBLIC REPORTING AND PAY-FOR-PERFORMANCE

In the current climate of increased attention for patient safety and cost containment, HAI rates are increasingly used for interfacility comparison, public reporting and pay-for-performance schemes[63,64]. These programs aim to achieve quality improvements by guiding (patient) choice of providers and stimulating hospitals to improve their performance[65-67]. The intended effect of these programs depends on two key premises. The events targeted by the programs should be preventable through the introduction of best-practice prevention strategies and, ideally, all measured variation between centers results from true difference in quality of care. The degree of preventability of the events currently included in public reporting programs is debated; defining 'never-events' and setting zero-targets has been argued by some, but opposed by others[68-72]. Secondly, factors other than quality of care also determine the variation in HAI rates across facilities. Noise

may not only be introduced by differences in surveillance methodology but also by differences in the characteristics of the underlying patient population[39,73].

Correcting for underlying patient characteristics

A prerequisite to drawing valid inferences on differences in quality of care is the availability of practical and accurate methods of case-mix adjustment[45,63,73,74]. Risk-adjustment models for surgical site infections have been refined in recent years, but future improvements are needed to ensure that these only include *a priori* risk factors and do not correct for intermediate factors also reflecting quality of care (e.g., duration of surgery). For other types of HAI, risk-adjustment is currently fairly crude and limited to stratification on hospital or patient characteristics. No automated systems that employ routine care data are currently available on a large scale, although their development and implementation could be coupled to efforts necessary for automated HAI surveillance[75-80]. In this thesis, we provide a first step towards employing clinical routine care data to develop a severity of disease adjustment model for mechanically ventilated patients (Ch. 7). This model pre-emptively considered possible vagaries introduced in routine care data as a result of differences in clinician behavior or data collection practices and aimed to identify a method that has the most optimal trade-off between feasibility and precision. The results of this study are promising though more work is necessary to assess the robustness of the model in different settings and develop methods that relate HAI incidence to the underlying risk profile.

Rates of HAI, after case-mix adjustment, can be presented in several ways to assess healthcare quality. Calculating standardized infection ratios with confidence intervals allows for identification of outlying hospitals[81,82] and ranking of hospitals has also been used to assess surveillance methods and measure relative performance[18,47,83]. Ranking, however, suffers from important conceptual limitations as this does not reflect the actual size of the difference between hospitals and the intercorrelation between positions may lead to unstable ranks, in particular for hospitals with average performance[84]. Random variation between (small) hospitals further limits the validity of ranking[85] and the use of random-effects models to account for natural variation across hospitals has been advocated[86]. Control charts and funnel plots may facilitate visual interpretation of differences between hospitals or over time[87,88].

Competing outcomes as an additional dimension

A conceptual dimension of HAI surveillance that is commonly overlooked in reporting of infection rates is the occurrence of competing (adverse) events. These competing events

can be illustrated by the ventilator-associated events (VAE) algorithm. In this framework, all patients that receive two or more days of mechanical ventilation are considered at risk and included in the denominator, but only those that were mechanically ventilated for at least four days are eligible for an event. As early death on the ventilator precludes the development of a VAE, it can be considered a competing outcome. For VAE, and in fact all types of HAI, death of the patient ends the time at risk and thus lowers absolute risk of developing an infection. In the case of failure-to-rescue of a severely ill patient, a higher proportion of early deaths will reduce the incidence of VAE but does not reflect adequate quality of care. Mortality, the most severe adverse event, should therefore be included in the interpretation of HAI incidence. This interdependence, however, is often ignored.

In the case of VAE, an additional (beneficial) competing outcome is early discontinuation of mechanical ventilation as, by definition, patients successfully weaned from the ventilator within four days cannot develop a VAE. The relative incidences of the competing outcomes (VAE, early death and early discharge) are dependent both on the baseline severity of disease of the patient and the quality of care provided. In fact, linking predicted baseline severity of disease to VAE incidence revealed that the observed correlation is largely mediated by accurately predicting the proportion of patients remaining on the ventilator long enough to be eligible for VAE, or in other words the competing events process (Ch. 7). Ideally, a quality measure and associated risk-adjustment method should account for this competing events framework. Restriction of the denominator to only those patients receiving prolonged mechanical ventilation does not solve this problem as it introduces a form of immortal-time bias and penalizes hospitals that either achieve rapid discontinuation of ventilation in less severely ill patients or manage to keep high-risk patient alive for more than four days. And, as illustrated above, the unit of measure chosen as the denominator will also affect the rates observed. Future work will need to identify the most accurate method of incorporating these competing events processes in quality measures, and develop a method of feedback that has an intuitive interpretation to the stakeholders involved.

Automated surveillance in the framework of quality evaluation

Within Donabedian's framework of outcome-, process- and structure measures[89], the debate on the relative value of process and outcome measures for quality improvement has not yet been closed. Public reporting and payment programs currently target a combination of process and outcome measures[16,90,91]. Clinical outcomes are more difficult to measure but may reflect relevant patient outcomes that motivate change, whereas process indicators may identify additional room for improvement and provide rapidly actionable data even when patient outcomes are already favorable or adverse

events are rare. Process indicators have, however, been criticized for their uncertain correlation with patient outcomes[92-95]. Automation of HAI surveillance may contribute to increased adoption of outcome measures by improving the efficiency and consistency of identification of (adverse) patient outcomes, although as elaborated upon previously, their validity is not guaranteed by the transition to automated systems. Importantly, more and more process measures are collected electronically from medical records and many of the concerns regarding variability in data generation and extraction also apply to their interpretation[43].

(Un)intended consequences of public reporting and pay-for-performance

Several studies have assessed the effect of public reporting and pay-for-performance programs on the incidence of HAI. Most studies have not found a measurable improvement in HAI rates resulting from non-payment programs, although some found increased attention to process of care and quality initiatives and perhaps benefits will accrue in the future[66,96-99]. Concurrently, introduction of financial incentives evoked warnings for unintended – and undesirable – consequences that may be deleterious for both the care patients receive and the utility of surveillance data to drive patient-safety improvement[91,100].

Measuring events that are not preventable but reflect underlying disease may encourage hospitals to resort to cherry-picking and threaten equitable access to healthcare. In addition, the focus of such programs on a limited selection of HAI may promote tunnel vision, thereby overlooking other safety needs[91]. For example, although critically ill patients benefit from surveillance efforts due to their high incidence of HAI, the greatest absolute numbers of HAI occur in patients not admitted to intensive care units. In addition, a recent survey showed that about 50% of HAI are not linked to specific procedures or devices and are currently not targeted by most (mandatory) surveillance programs[101,102].

Secondly, incorporating HAI rates in public reporting schemes or coupling them to financial incentives may put the reliability of surveillance at risk, thereby undermining the potential of surveillance programs to drive quality improvement and unjustly punishing hospitals that invest in thorough, high-quality surveillance programs[71,103]. An unexplained decrease in HAI rates measured by administrative data compared to trends in clinical surveillance data suggests changes in coding practices, or maybe even gaming, as a result of these programs[12,104,105]. Manual surveillance methods and clinical outcome

measures may also be affected by these financial incentives, for example by stricter application of case-definitions[39,106]. A recently published recommendation from the Healthcare Infection Control Practices Advisory Committee (HICPAC) on surveillance in the context of public reporting aims to provide guidance on preventing this collateral damage, for example by strongly advising against clinical adjudication or veto when determining HAI status[45]. Whether such guidance will be sufficient, however, may be questioned.

Can automated HAI surveillance systems provide relief to these unintended consequences of public reporting, described by some as a 'minefield in a quagmire'[107]? Undeniably, introduction of automated surveillance has the potential to increase the capacity of surveillance efforts and thereby reduce tunnel vision. Also, addressing the effort-dependent nature of case-finding and automated application of definitions will improve the reliability of rates obtained[4,37]. However, although automated surveillance systems help to generate more data in a consistent matter, many (unwanted) sources of variability remain. Adoption of automated systems will not in itself ensure comparability of HAI rates and does not guarantee data with a meaningful interpretation. Moreover, the ongoing transition to objective case-definitions that are amenable to automated implementation requires judicious evaluation to assure the generation of actionable data that can contribute to improved patient care.

CONCLUDING REMARKS

Automated surveillance of HAI using routine care data can improve the capacity and quality of surveillance by consistently identifying patients at high risk of HAI. The intended application of automated systems is a key determinant of the required performance characteristics of novel methods. Future research expanding the scope of automated surveillance systems, diligent implementation and ongoing rigorous validation studies of the novel methods will assist in finding the optimal data sources and striking the balance between simple and complex methods of surveillance. Development of adequate methods to correct for differences in severity of underlying disease and incorporation of competing outcomes will be necessary to perform valid comparisons of HAI incidence across hospitals.

REFERENCE LIST

1. Safdar N, Anderson DJ, Braun BI et al. The evolving landscape of healthcare-associated infections: recent advances in prevention and a road map for research. *Infect Control Hosp Epidemiol.* **2014**;35:480-493.
2. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. *Clin Infect Dis.* **2009**;48:1268-1275.
3. Halpin H, Shortell SM, Milstein A, Vanneman M. Hospital adoption of automated surveillance technology and the implementation of infection prevention and control programs. *Am J Infect Control.* **2011**;39:270-276.
4. Trick WE. Decision making during healthcare-associated infection surveillance: a rationale for automation. *Clin Infect Dis.* **2013**;57:434-440.
5. Niedner MF. The harder you look, the more you find: Catheter-associated bloodstream infection surveillance variability. *Am J Infect Control.* **2010**;38:585-595.
6. Stevens JP, Kachniarz B, Wright SB et al. When Policy Gets It Right: Variability in U.S. Hospitals' Diagnosis of Ventilator-Associated Pneumonia. *Crit Care Med.* **2013**;42:497-503.
7. Dixon-Woods M, Leslie M, Bion J, Tarrant C. What counts? An ethnographic study of infection data reported to a patient safety program. *Milbank Q.* **2012**;90:548-591.
8. Woeltje KF. Moving into the future: electronic surveillance for healthcare-associated infections. *J Hosp Infect.* **2013**;84:103-105.
9. Freeman R, Moore LS, Garcia AL, Charlett A, Holmes A. Advances in electronic surveillance for healthcare-associated infections in the 21st Century: a systematic review. *J Hosp Infect.* **2013**;84:106-119.
10. De Bruin JS, Seeling W, Schuh C. Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review. *J Am Med Inform Assoc.* **2014**;In Press.
11. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care.* **2003**;12 Suppl 2:ii58-ii63.
12. Farmer SA, Black B, Bonow RO. Tension between quality measurement, public quality reporting, and pay for performance. *JAMA.* **2013**;309:349-350.
13. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of Administrative Code Data for the Surveillance of Healthcare-Associated Infections: A Systematic Review and Meta-Analysis. *Clin Infect Dis.* **2013**.
14. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. *Clin Infect Dis.* **2009**;49:949-955.
15. Centers for Medicare and Medicaid Services. Hospital-Acquired Conditions. **2012**. http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html. Last accessed: Nov 25 2013.
16. National Quality Forum. Quality Positioning System. **2014**. http://www.qualityforum.org/Qps/QpsTool.aspx. Last accessed: May 6 2014.
17. Calderwood MS, Ma A, Khan YM et al. Use of Medicare Diagnosis and Procedure Codes to Improve Detection of Surgical Site Infections following Hip Arthroplasty, Knee Arthroplasty, and Vascular Surgery. *Infect Control Hosp Epidemiol.* **2012**;33:40-49.
18. Huang SS, Placzek H, Livingston J et al. Use of Medicare Claims to Rank Hospitals by Surgical Site Infection Risk following Coronary Artery Bypass Graft Surgery. *Infect Control Hosp Epidemiol.* **2011**;32:775-783.
19. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of Administrative Code Data for the Surveillance of Healthcare-Associated Infections: A Systematic Review and Meta-Analysis. *Clin Infect Dis.* **2014**;58:688-696.
20. Trick WE, Zagorski BM, Tokars JI et al. Computer algorithms to detect bloodstream infections. *Emerg Infect Dis.* **2004**;10:1612-1620.
21. Woeltje KF, McMullen KM, Butler AM, Goris AJ, Doherty JA. Electronic surveillance for healthcare-associated central line-associated bloodstream infections outside the intensive care unit. *Infect Control Hosp Epidemiol.* **2011**;32:1086-1090.

22. Brossette SE, Hacek DM, Gavin PJ et al. A laboratory-based, hospital-wide, electronic marker for nosocomial infection: the future of infection control surveillance? *Am J Clin Pathol.* **2006**;125:34-39.
23. Leth RA, Norgaard M, Uldbjerg N, Thomsen RW, Moller JK. Surveillance of selected post-caesarean infections based on electronic registries: validation study including post-discharge infections. *J Hosp Infect.* **2010**;75:200-204.
24. Yokoe DS, Khan Y, Olsen MA et al. Enhanced surgical site infection surveillance following hysterectomy, vascular, and colorectal surgery. *Infect Control Hosp Epidemiol.* **2012**;33:768-773.
25. Knepper BC, Young H, Jenkins TC, Price CS. Time-saving impact of an algorithm to identify potential surgical site infections. *Infect Control Hosp Epidemiol.* **2013**;34:1094-1098.
26. Magill SS, Klompas M, Balk R et al. Developing a new, national approach to surveillance for ventilator-associated events*. *Crit Care Med.* **2013**;41:2467-2475.
27. Agency for Healthcare Quality and Research. National Quality Forum (NQF) Endorsed Individual and Composite Measures. **2013**. http://www.qualityindicators.ahrq.gov/Downloads/Modules/V45/Module_NQF_Endorsement_V4.5.pdf. Last accessed: Mar 20 2014.
28. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J Med.* **2010**;363:501-504.
29. Wisniewski MF, Kieszkowski P, Zagorski BM, Trick WE, Sommers M, Weinstein RA. Development of a clinical data warehouse for hospital infection control. *J Am Med Inform Assoc.* **2003**;10:454-462.
30. NFU: Nederlandse Federatie van Universitair Medische Centra. Registratie aan de bron: visie op documentatie en gebruik van zorggegevens 2013 - 2020. **2013**
31. Clinical information Modeling Initiative (CIMI). **2014**. http://informatics.mayo.edu/CIMI/index.php/Main_Page. Last accessed:
32. Borlawsky T, Hota B, Lin MY et al. Development of a reference information model and knowledgebase for electronic bloodstream infection detection. *AMIA Annu Symp Proc.* **2008**;56-60.
33. Joukes, E. Measuring Quality of Care by Reusing Routinely Collected Data (MSc Thesis). **2013**
34. van Mourik MS, Troelstra A, van Solinge WW, Moons KG, Bonten MJ. Automated surveillance for healthcare-associated infections: opportunities for improvement. *Clin Infect Dis.* **2013**;57:85-93.
35. De Bruin JS, Adlassnig KP, Blacky A, Mandl H, Fehre K, Koller W. Effectiveness of an automated surveillance system for intensive care unit-acquired infections. *J Am Med Inform Assoc.* **2013**;20:369-372.
36. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control.* **2008**;36:309-332.
37. Lin MY, Hota B, Khan YM et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. *JAMA.* **2010**;304:2035-2041.
38. Oostenbrink R, Moons KG, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: prospects and problems. *J Clin Epidemiol.* **2003**;56:501-506.
39. Klompas M. Eight initiatives that misleadingly lower ventilator-associated pneumonia rates. *Am J Infect Control.* **2012**;40:408-410.
40. Boyce JM, Nadeau J, Dumigan D et al. Obtaining blood cultures by venipuncture versus from central lines: impact on blood culture contamination rates and potential effect on central line-associated bloodstream infection reporting. *Infect Control Hosp Epidemiol.* **2013**;34:1042-1047.
41. Janssen KJ, Vergouwe Y, Donders AR et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem.* **2009**;55:994-1001.
42. Bosman RJ, Oudemane van Straaten HM, Zandstra DF. The use of intensive care information systems alters outcome prediction. *Intensive Care Med.* **1998**;24:953-958.
43. Kern LM, Malhotra S, Barron Y et al. Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. *Ann Intern Med.* **2013**;158:77-83.
44. Burstin H. The journey to electronic performance measurement. *Ann Intern Med.* **2013**;158:131-132.
45. Talbot TR, Bratzler DW, Carrico RM et al. Public reporting of health care-associated surveillance data: recommendations from the healthcare infection control practices advisory committee. *Ann Intern Med.* **2013**;159:631-635.

46. Ming DY, Chen LF, Miller BA, Anderson DJ. The impact of depth of infection and postdischarge surveillance on rate of surgical-site infections in a network of community hospitals. *Infect Control Hosp Epidemiol.* **2012**;33:276-282.
47. Yokoe DS, Avery TR, Platt R, Huang SS. Reporting Surgical Site Infections Following Total Hip and Knee Arthroplasty: Impact of Limiting Surveillance to the Operative Hospital. *Clin Infect Dis.* **2013**.
48. Lankiewicz JD, Yokoe DS, Olsen MA et al. Beyond 30 Days: Does Limiting the Duration of Surgical Site Infection Follow-up Limit Detection? *Infect Control Hosp Epidemiol.* **2012**;33:202-204.
49. Dicks KV, Lewis SS, Durkin MJ et al. Surveying the surveillance: surgical site infections excluded by the january 2013 updated surveillance definitions. *Infect Control Hosp Epidemiol.* **2014**;35:570-573.
50. Mannien J, Wille JC, Snoeren RL, van den HS. Impact of postdischarge surveillance on surgical site infection rates for several surgical procedures: results from the nosocomial surveillance network in The Netherlands. *Infect Control Hosp Epidemiol.* **2006**;27:809-816.
51. Birnbaum D, Jarvis WR, Pronovost P, Zarate R. Statewide one-day survey of central lines: impact of different denominators on public reporting of infection rates. *Clinical Governance: An International Journal.* **2012**;17:217-222.
52. Wright MO, Kharasch M, Beaumont JL, Peterson LR, Robicsek A. Reporting catheter-associated urinary tract infections: denominator matters. *Infect Control Hosp Epidemiol.* **2011**;32:635-640.
53. Trick WE, Samore M. Denominators for device infections: who and how to count. *Infect Control Hosp Epidemiol.* **2011**;32:641-643.
54. Fakih MG, Todd GM, Kennedy EH et al. Introducing a population-based outcome measure to evaluate the effect of interventions to reduce catheter-associated urinary tract infection. *Am J Infect Control.* **2011**;40:359-364.
55. Naaktgeboren CA, de Groot JA, van SM, Moons KG, Reitsma JB. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med.* **2013**;159:195-202.
56. de Groot JA, Bossuyt PM, Reitsma JB et al. Verification problems in diagnostic accuracy studies: consequences and solutions. *BMJ.* **2011**;343:d4770.
57. Klompas M, Platt R. Ventilator-associated pneumonia--the wrong quality measure for benchmarking. *Ann Intern Med.* **2007**;147:803-805.
58. Uckay I, Ahmed QA, Sax H, Pittet D. Ventilator-associated pneumonia as a quality indicator for patient safety? *Clin Infect Dis.* **2008**;46:557-563.
59. Wunderink RG. Ventilator-associated Complications, Ventilator-associated Pneumonia, and Newton's Third Law of Mechanics. *Am J Respir Crit Care Med.* **2014**;189:882-883.
60. Lilly CM, Ellison RT, III. Quality measures for critically ill patients: where does ventilator-associated condition fit in? *Chest.* **2013**;144:1429-1430.
61. Lilly CM, Landry KE, Sood RN et al. Prevalence and Test Characteristics of National Health Safety Network Ventilator-Associated Events. *Crit Care Med.* **2014**.
62. Centers for Disease Control and Prevention. Surveillance for Ventilator-Associated Events. **2013**. http://www.cdc.gov/nhsn/acute-care-hospital/vae/. Last accessed: Nov 26 2013.
63. Panzer RJ, Gitomer RS, Greene WH, Webster PR, Landry KR, Riccobono CA. Increasing demands for quality measurement. *JAMA.* **2013**;310:1971-1980.
64. Haustein T, Gastmeier P, Holmes A et al. Use of benchmarking and public reporting for infection control in four high-income countries. *Lancet Infect Dis.* **2011**;11:471-481.
65. Ketelaar NA, Faber MJ, Flottorp S, Rygh LH, Deane KH, Eccles MP. Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. *Cochrane Database Syst Rev.* **2011**;CD004538.
66. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med.* **2008**;148:111-123.
67. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care.* **2003**;41:I30-I38.
68. Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infect Control Hosp Epidemiol.* **2011**;32:101-114.

69. Bonten MJ. Healthcare epidemiology: Ventilator-associated pneumonia: preventing the inevitable. *Clin Infect Dis.* **2011**;52:115-121.
70. Brown J, Doloresco IF, Mylotte JM. "Never events": not every hospital-acquired infection is preventable. *Clin Infect Dis.* **2009**;49:743-746.
71. Dixon-Woods M, Perencevich EN. When counting central line infections counts. *Infect Control Hosp Epidemiol.* **2013**;34:555-557.
72. Lambert ML, Silversmit G, Savey A et al. Preventable proportion of severe infections acquired in intensive care units: case-mix adjusted estimations from patient-based surveillance data. *Infect Control Hosp Epidemiol.* **2014**;35:494-501.
73. Harris AD, McGregor JC. The importance of case-mix adjustment for infection rates and the need for more research. *Infect Control Hosp Epidemiol.* **2008**;29:693-694.
74. Bratzler DW. Accountability for Surgical Site Infections: Will the Playing Field Be Level? *Clin Infect Dis.* **2013**.
75. Dudeck MA, Weiner LM, len-Bridson K et al. National Healthcare Safety Network (NHSN) report, data summary for 2012, Device-associated module. *Am J Infect Control.* **2013**;41:1148-1166.
76. Rijksinstituut voor Volksgezondheid en Mileu. PREZIES: PREventie van ZIEkenhuisinfections door Surveillance. **2013**. http://www.rivm.nl/Onderwerpen/P/PREZIES. Last accessed: Mar 15 2014.
77. Berrios-Torres SI, Mu Y, Edwards JR, Horan TC, Fridkin SK. Improved risk adjustment in public reporting: coronary artery bypass graft surgical site infections. *Infect Control Hosp Epidemiol.* **2012**;33:463-469.
78. Mu Y, Edwards JR, Horan TC, Berrios-Torres SI, Fridkin SK. Improving risk-adjusted measures of surgical site infection for the national healthcare safety network. *Infect Control Hosp Epidemiol.* **2011**;32:970-986.
79. Fukuda H, Morikane K, Kuroki M et al. Toward the rational use of standardized infection ratios to benchmark surgical site infections. *Am J Infect Control.* **2013**;41:810-814.
80. Roy MC, Herwaldt LA, Embrey R, Kuhns K, Wenzel RP, Perl TM. Does the Centers for Disease Control's NNIS system risk index stratify patients undergoing cardiothoracic operations by their risk of surgical-site infection? *Infect Control Hosp Epidemiol.* **2000**;21:186-190.
81. Rioux C, Grandbastien B, Astagneau P. The standardized incidence ratio as a reliable tool for surgical site infection surveillance. *Infect Control Hosp Epidemiol.* **2006**;27:817-824.
82. Malpiedi, P. J., Peterson, K. D., Soe, M. M., Edwards, J. R., Scott, R. D., Wise, M. E., Baggs, J, Yi, S. H., Dudeck, M. A., Arnold, K. E., Weiner, L. M., Rebmann, C. A., Srinivasan, A., Fridkin, S., and McDonald, L. C. 2011 National and State Healthcare-Associated Infections Standardized Infection Ratio Report. **2013**
83. Birnbaum D, Zarate R, Marfin A. SIR, you've led me astray! *Infect Control Hosp Epidemiol.* **2011**;32:276-282.
84. Siregar S, Groenwold RH, Jansen EK, Bots ML, van der GY, van Herwerden LA. Limitations of ranking lists based on cardiac surgery mortality rates. *Circ Cardiovasc Qual Outcomes.* **2012**;5:403-409.
85. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. *BMJ Qual Saf.* **2011**;20:869-874.
86. van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. *Br J Surg.* **2013**.
87. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med.* **2005**;24:1185-1202.
88. Lawson EH, Hall BL, Esnaola NF, Ko CY. Identifying worsening surgical site infection performance: control charts versus risk-adjusted rate outlier status. *Am J Med Qual.* **2012**;27:391-397.
89. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q.* **1966**;44:Suppl-206.
90. Health Council of the Netherlands (Gezondheidsraad). Publieke indicatoren voor de kwaliteit van curatieve zorg: De stand van de discussie. **2013**
91. Mannion R, Davies HT. Payment for performance in health care. *BMJ.* **2008**;336:306-308.
92. Bonten MJ, Friedrich A, Kluytmans JA, Vandenbroucke-Grauls CM, Voss A, Vos MC. [Infection prevention in Dutch hospitals; results say more than process indicators]. *Ned Tijdschr Geneeskd.* **2014**;158:A7395.
93. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care.* **2001**;13:475-480.

94. Siregar S, Groenwold RH, Versteegh M, van Herwerden LA. [Performance indicators for hospitals]. *Ned Tijdschr Geneeskd.* **2012**;156:A5487.

95. Rodriguez-Bano, J and Gastmeier, P. ECCMID 2014 pro-con debate: Infections or proceducers. **2014**

96. Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of premier pay for performance on patient outcomes. *N Engl J Med.* **2012**;366:1606-1615.

97. Lee GM, Kleinman K, Soumerai SB et al. Effect of nonpayment for preventable infections in U.S. hospitals. *N Engl J Med.* **2012**;367:1428-1437.

98. Pakyz AL, Edmond MB. Influence of state laws mandating reporting of healthcare-associated infections: the case of central line-associated bloodstream infections. *Infect Control Hosp Epidemiol.* **2013**;34:780-784.

99. Schuller K, Probst J, Hardin J, Bennett K, Martin A. Initial impact of Medicare's nonpayment policy on catheter-associated urinary tract infections by hospital characteristics. *Health Policy.* **2013**.

100. Hartmann CW, Hoff T, Palmer JA, Wroe P, Dutta-Linn MM, Lee G. The Medicare policy of payment adjustment for health care-associated infections: perspectives on potential unintended consequences. *Med Care Res Rev.* **2012**;69:45-61.

101. Magill SS, Edwards JR, Bamberg W et al. Multistate point-prevalence survey of health care-associated infections. *N Engl J Med.* **2014**;370:1198-1208.

102. Lewis SS, Moehring RW, Chen LF, Sexton DJ, Anderson DJ. Assessing the relative burden of hospital-acquired infections in a network of community hospitals. *Infect Control Hosp Epidemiol.* **2013**;34:1229-1230.

103. Haut ER, Pronovost PJ. Surveillance bias in outcomes reporting. *JAMA.* **2011**;305:2462-2463.

104. Black B. Emerald article: Public reporting of HAI rates: what we (mostly don't) know. *Clinical Governance: An International Journal.* **2012**;17:124-133.

105. Calderwood MS, Kleinman K, Soumerai SB et al. Impact of Medicare's Payment Policy on Mediastinitis Following Coronary Artery Bypass Graft Surgery in US Hospitals. *Infect Control Hosp Epidemiol.* **2014**;35:144-151.

106. Klompas M. What can we learn from international ventilator-associated pneumonia rates?*. *Crit Care Med.* **2012**;40:3303-3304.

107. McGuckin M, Govednik J, Hyman D, Black B. Public reporting of healthcare-associated infections: epidemiologists' perspectives. *Infect Control Hosp Epidemiol.* **2013**;34:1201-1203.

Closing pages

SUMMARY

Healthcare-associated infections (HAI) are among the most prevalent complications of medical care, affecting one in twenty-five hospitalized patients on any given day. Surveillance of HAI and calculation of reliable incidence rates is a key component of successful infection prevention programs. However, traditional surveillance by manual chart review is time-consuming and prone to error. Developments in healthcare information technology and the increasing demands for HAI surveillance have driven the development and implementation of automated surveillance systems. These systems employ routine care data stored in electronic health records to identify (probable) cases of HAI. The aim of the studies in this thesis is to further explore the possibilities of automated HAI surveillance using routine care data and compare different types of routine care data and approaches to automated surveillance.

Harnessing clinical routine care data

Chapter 2 presents the development of an automated surveillance system for a prototype HAI. The aim was to retrospectively detect drain-related meningitis (DRM), a common infection in neurosurgical patients receiving an external ventricular (EVD) or lumbar drain (ELD). In this retrospective cohort study, the outcome of routine surveillance was considered the reference standard, thus all patients receiving an EVD or ELD (2004 – 2009) were evaluated for the development of DRM through chart review and standardized diagnostic criteria. To identify patients with DRM using routinely collected data extracted from a clinical data warehouse, a multivariable prediction model was developed using logistic regression and bootstrapping techniques to adjust for possible overfitting. This model included the type and number of drains placed, blood leukocyte count, C-reactive protein, cerebrospinal fluid leukocyte count and culture result, number of antibiotics started during admission and initiation of empiric antibiotic therapy and could accurately detect cases of DRM. Discriminatory power was excellent (area under the ROC curve 0.97) and achieved 98.8% sensitivity at 56.9% positive predictive value. In addition, predicted yearly infection rates concurred with observed infection rates. These results supported the feasibility of automated HAI detection using multivariable prediction models.

In **chapter 3**, the aim was to perform a temporal validation and update of the automated approach developed in chapter 2 using the same reference standard and methods. For temporal validation, the original regression model was applied to data from 2010 – 2011; subsequently, the development and validation set were combined to update the model by determining whether extension with newly available predictors was necessary and to

optimally re-estimate coefficients. In model validation, all cases of DRM were detected by the model at 58.8% positive predictive value. The updated model incorporates Gram stain results as an additional predictor and after correction for possible overfitting the area under the ROC curve was 0.96 with adequate group-level prediction. At 98% sensitivity, the model had 60% positive predictive value for DRM and could decrease the burden for manual chart review by 75%. This initial validation study showed that automated surveillance using a multivariable prediction model maintains discriminatory power and calibration in an independent patient population; the updated model incorporates all available data and also had favorable performance characteristics.

Chapter 4 reviews existing automated surveillance systems and compares different approaches to automated HAI detection. Surveillance systems based on classification algorithms using microbiology results, antibiotic use and/or discharge codes have increased the efficiency and completeness of surveillance by pre-selecting high-risk patients for manual review. However, as illustrated by the data obtained from the DRM studies, shifting to automated surveillance using multivariable regression (prediction) models allowed for even more efficient detection of infection. Ongoing developments in healthcare information technology will make implementation of the latter surveillance systems increasingly feasible. Importantly, as with manual methods, several challenges do remain such as completeness of post-discharge surveillance and adequate adjustment for differences in underlying patient characteristics, especially for comparison of HAI rates across institutions.

In the study presented in **chapter 5**, the multivariable regression modeling approach for detection of drain-related meningitis is externally validated and its merits are compared to conventional methods of automated surveillance. In this prospective cohort study, three hospitals assessed the accuracy and efficiency of two automated surveillance methods - the previously developed multivariable regression model and a classification algorithm - for detecting DRM in patients receiving an ELD or EVD, again using surveillance by manual chart review as reference standard. In total, DRM occurred in 37 of 366 external cerebrospinal fluid drainage episodes. Although there were considerable differences between the hospitals regarding patient characteristics, diagnostic and treatment practices, the multivariable regression model had favorable performance characteristics. Discriminatory power was high (area under the ROC curve 0.91 – 1.00 by hospital) with adequate prediction of group-level incidence. High-risk patients requiring manual confirmation were identified with 97.3% sensitivity and 52.2% positive predictive value, decreasing the workload for manual surveillance by 81%. The multivariable regression approach was more efficient than classification algorithms in 2 of 3 hospitals. This study

thus provides support for the reproducibility and generalizability of the developed surveillance system.

Concerns with the reliability of traditional surveillance methods are among the drivers behind a movement towards the introduction of more objective case-definitions that are amenable to electronic implementation. **Chapter 6** presents the results of a study determining the validity of a novel surveillance paradigm for ventilator-associated events (VAE) that was recently introduced in the United States as a replacement of conventional surveillance for ventilator associated pneumonia (VAP). This novel algorithm identifies three nested entities: ventilator-associated conditions (VAC), infection-related VAC (IVAC) and VAP. The VAE algorithm was electronically implemented in two Dutch academic medical centers and compared to conventional, prospective VAP surveillance (2,296 admissions in 2,080 patients). The incidences of VAC, IVAC, VAE-VAP and VAP according to prospective surveillance were 10.0, 4.2, 3.2 and 8.0 per 1000 ventilation days, respectively. The VAE algorithm detected at most 32% of the VAP patients identified by prospective surveillance and positive predictive value of a VAE signal for VAP was mediocre. Assessment of clinical diagnoses underlying VAC signals revealed that these were most often caused by volume overload and infections, but not necessarily VAP. Attributable mortality was calculated for all types of events using competing risk methods. Subdistribution hazards for mortality were 3.9 for VAC, 2.5 for IVAC, 2.0 for VAE-VAP and 7.2 for VAP identified by prospective surveillance. Strikingly, in sensitivity analyses comparing minor differences in electronic implementation, episodes of VAC identified and their estimated attributable mortality varied considerably.

As HAI rates are increasingly used for benchmarking of hospitals and pay-for-performance incentives, methods to accurately and efficiently correct for differences in hospitals' case-mix are required. However, existing risk adjustment models for HAI tend either to be very simplistic or to require extensive manual data collection. **Chapter 7** presents the development of a disease severity prediction model based solely on data routinely available in electronic health records for the purpose of group-level risk adjustment in mechanically ventilated patients. This retrospective cohort study included 20,028 mechanically ventilated patients from a single tertiary medical center in the United States (2006 – 2012), of whom 3,027 deceased in hospital. Model predictors were extracted from electronic data repositories and evaluated *a-priori* for suitability for inclusion. Models of increasing complexity for in-hospital mortality were built using logistic regression to optimize the balance between model precision and ease of implementation. A simple model including demographic information, type of intensive care unit, time to intubation, blood culture sampling, eight common laboratory tests and surgical procedures achieved

an area under the ROC curve of 0.87 (95% CI 0.86 – 0.88) with adequate calibration. The predicted risk of mortality was associated with occurrence of events detected by the VAE surveillance paradigm. Hence, accurate prediction of mortality risk in ventilated patients using electronic, routine care data was feasible using simple models and, with future validation studies, these estimates of disease severity may be useful for adjustment of HAI rates.

Administrative data

Routine care data may either be obtained from clinical sources such as laboratory testing or medication use or may be extracted from administrative databases. In **chapter 8**, the accuracy of discharge coding for the detection of DRM was assessed within the cohort used for development of the automated systems (Ch 2 and 3). In the absence of a specific code for DRM, a broad range of diagnosis codes was considered as possibly reflective of DRM. This study showed that discharge coding data had low sensitivity (32%) and positive predictive value (35%) and therefore could not replace more complex surveillance systems using clinical routine care data. In addition, the discharge diagnoses and billing records had no incremental value within the models employing clinical routine care data to detect DRM.

Despite concerns regarding their accuracy, administrative data - including discharge diagnosis codes - are commonly used to determine the occurrence of HAI, either as external quality indicators or to support within-hospital surveillance programs. **Chapter 9** presents a systematic review assessing the accuracy of administrative data for surveillance of a broad range of HAI. The fifty-seven included studies underwent rigorous methodological quality assessment, and accuracy estimates were stratified by HAI type, risk of bias and intended application of the administrative data algorithm for internal or external quality evaluation. Study designs were very diverse with respect to the selection of administrative data and the definitions used to determine HAI presence. In addition, incomplete ascertainment of HAI status and lack of blinding was common, possibly introducing bias in the observed accuracy estimates. Observed sensitivity and positive predictive value were very variable and generally modest at best for all HAI types and no differences were observed between algorithms developed for internal or external use. Sensitivity was particularly low for device-associated infections that necessitate the identification of central lines or urinary catheters using administrative data. Given the limited, and highly variable, accuracy of administrative data for the detection of HAI observed in this study, their judicious use in the process of HAI surveillance is recommended. Future, robust, validation studies are imperative if administrative data continue to be relied upon for HAI surveillance.

Synthesis

Chapter 10 positions automated systems using routine care data in the current context of HAI surveillance. Automated surveillance of HAI using routine care data can improve the capacity and quality of surveillance by consistently identifying patients at high risk of HAI. Different types of routine care data and approaches to automated HAI surveillance are compared and recommendations are provided for developing novel systems. Although automated surveillance systems help to generate more data in a consistent matter, this by itself does not guarantee data with a meaningful interpretation and (unwanted) sources of variation may remain. In addition, the benefits and limitations of automated HAI surveillance in the context of outcome comparisons and (public) quality evaluation are discussed.

NEDERLANDSE SAMENVATTING

Naar schatting ontwikkelt één op de 25 ziekenhuispatiënten een infectie als gevolg van de medische behandeling die zij ondergaan. Hiermee zijn deze zogenaamde ziekenhuisinfecties, of zorginfecties, één van de meest voorkomende (deels) vermijdbare complicaties van zorg. Surveillance van zorginfecties door het systematisch registreren van het optreden van infecties en het terugkoppelen van incidentiecijfers aan de zorgverleners is een cruciaal onderdeel van succesvolle infectiepreventie programma's. De momenteel meest gangbare methode van infectieregistratie is handmatig dossieronderzoek van alle patiënten die in de risicopopulatie vallen – bijvoorbeeld omdat zij een operatie hebben ondergaan of een centrale lijn hebben gekregen. Deze handmatige registratie is arbeidsintensief en foutgevoelig; informatie moet uit veel verschillende bronnen bij elkaar worden gebracht voordat een gestandaardiseerde definitie kan worden toegepast. Ook laten deze definities soms ruimte voor subjectieve interpretatie, hetgeen een consistente identificatie van zorginfecties verder bemoeilijkt.

Deze beperkingen van huidige registratiemethoden in combinatie met de bredere invoering van elektronische patiëntendossiers binnen ziekenhuizen hebben de ontwikkeling van geautomatiseerde methoden om mogelijke zorginfecties te detecteren gestimuleerd. Deze surveillance systemen (her)gebruiken zorgdata opgeslagen in elektronische ziekenhuisinformatiesystemen om onderscheid te maken tussen patiënten met een laag en hoog risico op een doorgemaakte zorginfectie. Een dergelijke risicostratificatie zorgt voor een grotere capaciteit en een meer consistente surveillancemethodiek (hoofdstuk 1, introductie).

De studies in dit proefschrift onderzoeken de mogelijkheden van geautomatiseerde surveillance op basis van zorgdata. Verschillende methoden van geautomatiseerde surveillance en typen zorgdata worden vergeleken, de bruikbaarheid en robuustheid van een alternatieve definitie wordt onderzocht evenals de mogelijkheden om zorgdata te gebruiken in correctiemethoden voor verschillen in ernst van onderliggend lijden. Het proefschrift sluit af met een samenvattende discussie.

Klinische zorgdata

Onder klinisch zorgdata worden verstaan alle klinische gegevens die over een patiënt verzameld en vastgelegd worden tijdens het routine proces van zorg, bijvoorbeeld operatiegegevens, antibioticagebruik, uitslagen van het klinisch-chemisch laboratorium en de medische microbiologie. Deze gegevens worden geregistreerd in het elektronische ziekenhuisinformatiesysteem en kunnen worden ontsloten via een zogenaamd 'clinical

data warehouse'. Omdat de gegevens verzameld worden tijdens het routine zorgproces, is hun verzameling niet gestandaardiseerd maar afhankelijk van de conditie van de patiënt en de diagnostische en therapeutische voorkeuren van de arts – hier moet dan ook rekening mee worden gehouden wanneer dergelijke gegevens worden (her)gebruikt.

Hoofdstuk 2 beschrijft de ontwikkeling van een surveillance aanpak voor de detectie van drain-gerelateerde meningitis (DRM), een complicatie die relatief frequent optreedt bij het gebruik van externe liquor drainage systemen zoals externe ventrikeldrains (EVDs) of externe lumbaaldrains (ELDs) in neurochirurgische patiënten. De surveillance aanpak is ontwikkeld op 537 patiënten die in de periode 2004 tot 2009 een externe liquordrain hebben ontvangen. Als referentiestandaard gold de uitkomst van 'traditionele surveillance': het voorkomen van DRM zoals vastgesteld door middel van retrospectief dossieronderzoek met behulp van een gestandaardiseerde definitie. Om patiënten met een hoog risico op een doorgemaakte DRM te identificeren is een multivariabel regressiemodel ontwikkeld dat de volgende predictoren bevat: type liquordrain, aantal drainwissels, uitslagen van de liquorkweek, leukocyten aantal in liquor en bloed, CRP, aantal antibioticawissels en initiatie van empirische antimicrobiële therapie gericht tegen DRM. Dit systeem behaalde een hoog onderscheidend vermogen, ook na interne validatie en correctie voor over-optimisme (oppervlakte onder de ROC curve van 0.97). Op basis van een drempelwaarde kunnen hoog-risico patiënten worden geïdentificeerd met een sensitiviteit van 98.8%, bij een positieve voorspellende waarde voor het aantonen van meningitis van 56.9%. Hierdoor kan de werklast van handmatig statusonderzoek worden verminderd met 75 procent terwijl nagenoeg alle doorgemaakte episodes van DRM worden opgespoord. Daarnaast kan op basis van dit logistische regressie model een voorspelde incidentie op groepsniveau worden berekend zonder dat er statusonderzoek nodig is; deze voorspelde incidenties kwamen overeen met de geobserveerde incidenties.

In **hoofdstuk 3** wordt het in hoofdstuk 2 ontwikkelde model gevalideerd op een onafhankelijke patiëntenpopulatie (temporele validatie, 2010 - 2011) en wordt het model verder verfijnd door oorspronkelijke en nieuwe data te combineren (model update). Tijdens de validatiestudie zijn dezelfde methoden gebruikt om DRM vast te stellen en is het model uit hoofdstuk 2 ongewijzigd toegepast. Hiermee werden alle gevallen van DRM geïdentificeerd, met een positief voorspellende waarde van 58.8%. In het gereviseerde model worden ook de uitslagen van het Gram preparaat meegenomen en werd een vergelijkbare sensitiviteit en positief voorspellende waarde behaald. Deze initiële validatiestudie bevestigt de potentie van geautomatiseerde surveillance op basis van multivariabele regressiemodellen.

Hoofdstuk 4 geeft een overzicht van bestaande geautomatiseerde surveillance systemen en beschrijft voor- en nadelen van verschillende methoden. De meerderheid van systemen is gebaseerd op classificatiealgoritmen – op basis van bijvoorbeeld microbiologie uitslagen en antibioticagebruik worden patiënten in een serie van dichotome stappen geclassificeerd als laag of hoog risico, waarbij de laatste groep statusonderzoek moet ondergaan. Aan de hand van de data uit de DRM studies wordt geïllustreerd hoe de efficiëntie van systemen kan worden verbeterd door gebruik te maken van multivariabele regressiemodellen. Hoewel deze iets complexer zijn dan classificatiealgoritmen, bieden ze meer flexibiliteit en de mogelijkheid tot het direct schatten van de incidentie in op groepsniveau zonder de noodzaak voor handmatig statusonderzoek. Verdere ontwikkelingen in zorg-ICT, verbeterde registratie en ontsluiting van elektronische bronnen zullen implementatie van dergelijke systemen in toenemende mate mogelijk maken, al valt of staat het succes hiervan bij adequate registratie van gegevens aan de bron. Evenals met handmatige methoden liggen er nog uitdagingen in het verschiet, zoals het realiseren van surveillance na ontslag en correctie van incidentiecijfers voor verschillen in ernst van onderliggend lijden.

De geautomatiseerde surveillance methode voor DRM wordt verder gevalideerd in **hoofdstuk 5**. In drie ziekenhuizen wordt de accuratesse en efficiëntie van de geautomatiseerde aanpak vergeleken met traditionele surveillance en wordt de toegevoegde waarde van multivariabele regressiemodellen ten opzichte van classificatiealgoritmen bepaald. Deze prospectieve cohortstudie omvat 366 patiënten die een EVD of ELD hebben gekregen. Ondanks verschillen tussen de ziekenhuizen met betrekking tot de patiëntenpopulatie en diagnostische en therapeutische protocollen, was het onderscheidend vermogen van het multivariabele regressiemodel goed (oppervlakte onder de ROC curve 0.91 – 1.00 per ziekenhuis) met adequate schattingen van incidentie op groepsniveau. De sensitiviteit was 97.3%, met 52.2% positief voorspellende waarde en de werklast voor handmatig statusonderzoek kon met 80 procent verminderd worden. Het multivariabele regressiemodel was efficiënter dan het classificatiealgoritme in twee van de drie ziekenhuizen. Deze studie ondersteunt de reproduceerbaarheid en generaliseerbaarheid van de ontwikkelde surveillancemethodiek.

In **hoofdstuk 6** wordt een nieuw paradigma voor de surveillance van beademingsgerelateerde complicaties geëvalueerd. Deze aanpak is recent in de Verenigde Staten (V.S.) geïntroduceerd als alternatief voor conventionele surveillance voor beademingsgerelateerde pneumonie (ventilator-associated pneumonia, VAP). Diagnostiek van deze complicatie is notoir problematisch en heeft te lijden onder subjectieve interpretatie van definities die worden toegepast op een complexe patiëntenpopulatie. Het

nieuwe paradigma streeft naar meer objectieve en consistente surveillance door gebruik te maken van beademingsdrukken om mogelijke gevallen van beademingsgerelateerde complicaties (ventilator-associated conditions, VAC) op te sporen. In een getrapt algoritme wordt aan de hand van infectieparameters (objectieve klinische verschijnselen, antibiotica gebruik, microbiologie uitslagen) de diagnose verder versmald tot een infectie-gerelateerde beademingscomplicatie (IVAC) of een beademingsgerelateerde pneumonie (VAP). Naast objectiviteit was ook de mogelijkheid tot elektronische implementatie van het algoritme een belangrijk uitgangspunt bij het ontwikkelen van nieuwe definities. Uit deze retrospectieve cohort studie waarin 2296 beademingsepisoden zijn geanalyseerd blijkt dat de overeenstemming tussen de conventionele surveillance en de entiteiten in het nieuwe paradigma beperkt is (sensitiviteit en positief voorspellende waarde beide < 30%). Een post-hoc analyse laat zien dat een groot deel van de VAC signalen wordt veroorzaakt door volume overbelasting of infecties, maar niet noodzakelijkerwijs door VAP. De attributieve sterfte van de verschillende aandoeningen is berekend door middel van een competing-risks analyse, waarbij de subdistributie hazard ratio 3.9 was voor VAC, 2.5 voor IVAC, 2.0 voor VAP volgens het nieuwe algoritme en 7.2 voor VAP volgens conventionele surveillance. Dit ondersteunt de opvatting dat het nieuwe algoritme verschillende typen episodes van VAP identificeert. Opvallend was daarnaast dat hoewel elektronische implementatie van het nieuwe algoritme mogelijk was, subtiele verschillen in implementatie noemenswaardige gevolgen hebben voor de episodes van VAC die werden geïdentificeerd, evenals hun attributieve sterfte. Deze nieuwe – mogelijk meer objectieve – definities garanderen onder de huidige specificaties dus niet dat de uitkomsten van surveillance vergelijkbaar en consistent zijn.

Naast het identificeren van zorginfecties is voor betrouwbare interpretatie van incidentiecijfers ook een adequate correctie voor verschillen in ernst van onderliggend lijden (case-mix) nodig. Huidige correctie methoden voor zorginfecties zijn niet erg gedetailleerd of vergen uitgebreide handmatige dataverzameling. **Hoofdstuk 7** beschrijft de ontwikkeling van een risicostratificatiemodel volledig gebaseerd op elektronisch beschikbare gegevens. Deze retrospectieve cohortstudie omvat 20,028 beademde patiënten uit een Amerikaans derdelijns ziekenhuis, waarvan er 3,027 tijdens hun opname zijn overleden. Alle mogelijke databronnen werden *a priori* beoordeeld op (verwachte) haalbaarheid en betrouwbaarheid van elektronische data verzameling. Vervolgens zijn predictiemodellen van toenemende datacomplexiteit ontwikkeld om ziekenhuissterfte te voorspellen, waarbij de balans is opgezocht tussen precisie van de predictie en complexiteit van het model. Een betrekkelijk eenvoudig model met daarin type intensive care, demografische gegevens, tijd tot intubatie, afname van bloedkweken, acht veelvoorkomende laboratoriumbepalingen en operatiegegevens had een oppervlakte

onder de ROC curve van 0.87 (95% betrouwbaarheidsinterval 0.86 – 0.88) met adequate calibratie. Daarnaast was het voorspelde risico op sterfte geassocieerd met de kans dat een patiënt een beademingsgerelateerde complicatie zal ontwikkelen. Routine zorgdata kan dus gebruikt worden in relatief eenvoudige predictiemodellen om het risico op overlijden betrouwbaar te voorspellen. Met toekomstige validatiestudies zullen de schattingen van de ernst van onderliggend lijden verder verbeterd worden en kunnen methoden worden ontwikkeld om met deze schatting de incidentiecijfers van zorginfecties betrouwbaar te interpreteren

Administratieve gegevens

Naast klinische zorgdata kan ook gebruik worden gemaakt van administratieve gegevens. In **hoofdstuk 8**, wordt de accuratesse van gecodeerde ontslagdiagnoses en financiële gegevens voor de detectie van DRM bepaald in dezelfde patiëntenpopulatie waarop het predictiemodel voor DRM is ontwikkeld (H2, H3). Ondanks een brede selectie van diagnosecodes die een mogelijke episode van DRM kunnen signaleren, was de sensitiviteit voor detectie van DRM slechts 32% met 33% positief voorspellende waarde. Daarnaast hadden diagnosecodes geen toegevoegde waarde ten opzichte van de surveillance methoden op basis van klinische zorgdata. In de Nederlandse setting lijken diagnosecodes voor DRM dus onvoldoende betrouwbaar voor surveillance.

Ondanks alom beschreven beperkingen van administratieve gegevens voor de surveillance van zorginfecties, worden zij wereldwijd nog veel gebruikt om kwaliteit van zorg en het optreden van zorginfecties te evalueren, dan wel als losstaande kwaliteitsindicator, dan wel binnen een breder geautomatiseerd surveillance systeem. **Hoofdstuk 9** presenteert de resultaten van een systematische review naar de diagnostische waarde van administratieve gegevens voor surveillance van het gehele spectrum aan zorginfecties. Uit een rigoureuze evaluatie van methodologische kwaliteit bleek dat deze bij 1 op de 3 van de zevenenvijftig geanalyseerde studies problematisch was, bijvoorbeeld omdat niet alle patiënten een (zelfde) referentiestandaard hebben ondergaan of omdat de observatoren niet geblindeerd waren. Ook ten aanzien van de gekozen methode als referentiestandaard, specificatie van de administratieve gegevens en patiëntenpopulaties waren de studies verschillend; om deze reden zijn er geen gepoolde analyses uitgevoerd. Over het algemeen was de accuratesse van administratieve gegevens voor detectie van zorginfecties ontoereikend, in het bijzonder voor infecties gerelateerd aan een centrale lijn of urinekatheter. Er waren geen noemenswaardige verschillen tussen algoritmen ontwikkeld als losstaande kwaliteitsindicator of gebruikt binnen een breder geautomatiseerd surveillance systeem.

In **hoofdstuk 10** worden de resultaten van de verschillende studies samengevat en in context geplaatst. De verschillende typen zorgdata en methoden van geautomatiseerde surveillance worden besproken, en hiaten in bestaande systemen worden toegelicht, samen met aanbevelingen voor de toekomst. Geautomatiseerde surveillance systemen kunnen de capaciteit en consistentie van surveillance vergroten, maar deze systemen op zich garanderen niet dat variabiliteit in surveillance methoden in zijn geheel is ondervangen. Daarnaast wordt ook ingegaan op verschuivende definities ten behoeve van (meer objectieve) geautomatiseerde surveillance en de noodzaak van adequate correctie voor verschillen in onderliggend lijden. Tot slot worden de gevolgen van publieke kwaliteitsindicatoren en prestatiebekostiging op betrouwbaarheid van surveillance van zorginfecties besproken en de mogelijke rol van geautomatiseerde surveillance hierin.

DANKWOORD

Hier ligt het, mijn proefschrift is af! Promoveren is geen solovoorstelling en er zijn velen die aan de totstandkoming van dit proefschrift hebben bijgedragen: mijn dank is groot. Een aantal mensen wil ik graag in het bijzonder bedanken.

Geachte prof.dr. M.J.M. Bonten, geachte promotor, beste Marc, toen ik als studente tegenover je zat op zoek naar een wetenschapsproject had jij nog wel een uitdaging liggen: veel elektronische data en de wens om surveillance van ziekenhuisinfecties te automatiseren. Daar is uiteindelijk dit proefschrift uit gekomen. Dank voor de steun en kansen die je mij tijdens dit promotietraject hebt gegeven en je vertrouwen in een goede afloop. De vrijheid om zelf koers te zetten in combinatie met een altijd bereikbaar kompas om bij te sturen was en is een prachtige springplank.

Geachte prof.dr. K.G.M. Moons, geachte promotor, beste Carl, iets later schoof je aan in mijn begeleidingsteam omdat predictie de "core business" bleek te worden. Met je scherpe oog voor methodologie en relevantie van de grote lijnen leg jij altijd de vinger op de zwakke plek. Dank voor je constructief kritische blik waarmee je er net nog iets meer uithaalde.

Dr. A. Troelstra, beste Annet, als co-promotor en hoofd van de ziekenhuishygiëne was jij vanaf het eerste uur bij dit onderzoek betrokken. Gedurende het hele traject kon ik voor advies bij je terecht, ook als het niet over de inhoud ging. Dank hiervoor!

Geachte leden van de beoordelingscommissie, prof.dr. M.M.E. Schneider, prof.dr. C.J. Kalkman, prof.dr. J.A.J.W. Kluytmans, prof.dr. K.C.B. Roes, prof.dr. M.C. Vos, dank dat u de tijd heeft genomen om mijn proefschrift te beoordelen.

Dr. Michael Klompas, dear Mike, thank you for your warm welcome at DPM in Boston. It is a pleasure to work with you, and I have brought back an invaluable amount of new knowledge and experience. Our meetings were the perfect mix of science, discussion and general wisdom. I sincerely hope we will be able to continue our trans-Atlantic collaboration.

Dr. R.H.H. Groenwold, beste Rolf, onder jouw begeleiding is de basis van dit onderzoek gelegd. Dank dat je mij als student de goede kant op stuurde, en ook daarna kon ik met korte en lange(re) methodologische vraagstukken altijd bij je terecht.

Beste afdeling ziekenhuishygiëne, beste Ada, Lia, Hetty, Marischka, Sylvia, Ben, Cathalijne, Kees, Roel, Sandra, Stephanie, zonder jullie dataverzameling was dit onderzoek niet gelukt! Dank voor het inkijkje in jullie werk en de waardevolle input. Ik kijk er naar uit om samen met jullie dit onderzoek in de praktijk te brengen.

Dataverzameling – ook uit elektronische databases – is geen sinecure. Zij die hebben bijgedragen aan de zoektocht naar data weten hier alles van: Hanneke den Breeijen, Leon Stijvers, Leslie Beks, Olaf Cremer, Ed Mulder, Wietze Pasma. Geachte prof.dr. Wouter van Solinge, dank dat u het gebruik van de UPOD voor deze studies steeds heeft ondersteund. Daarnaast waren vele studies niet mogelijk geweest zonder de inspanning van de afdeling neurochirurgie, in het bijzonder Jan Willem Berkelbach van der Sprenkel, en de Intensive Care.

Zonder de belangeloze inzet van velen in het Medisch Centrum Haaglanden (in het bijzonder Paul Dennesen, Mark Arts, arts-assistenten neurochirurgie en IC en de functioneel beheerders Fabio Bruna en Elly Kelleter-Gramsma) en het St. Elisabeth Ziekenhuis in Tilburg (o.a., Jolande Nelson, Piet Vos, Yvette Groenleer) was de multicenter validatie studie niet van de grond gekomen.

Geachte selectiecommissie van het Alexander Suermann programma, dank voor het vertrouwen dat u in dit onderzoek heeft gegeven. AS-buddies, aka de Su(p)ermannen, – Amani, Kim, Laura, Lotte, Margareta, Okan, Sylwia – jullie zijn een waardevol klankbord en bron van inspiratie! Onder bezielende leiding van Lucette hebben we mooie masterclasses en intervisies achter de rug.

Dear all at DPM and BWH, including Richard Platt, Meghan, Debbie, Grace, Michael C, Michael M, Julie, Sarah, Christina, Crystal, Liz and the team of infection preventionists, thank you for welcoming me and sharing your knowledge and fun.

Kamer G04.560, jullie zijn en waren mijn thuis op de afdeling! Evelien B, Daphne, Annemarie, Anne, Anna, Nienke de J, Sue, Tamara, Ron, wat gezelligheid in ons pijpenlaatje. Voornamelijk dames, veel chocola en een gezonde portie lachen is het recept voor een prachtige productieve tijd.

Lieve epi's, vele woensdagochtenden zaten wij al vroeg paraat om onze logistiek, dwarsliggers, regressiemodellen en confounders te bespreken. Met wie anders dan Bastiaan, David, Douwe, Evelien O, Fleur, Gerrita, Henok, Henri, Joppe, Jos, Lennie, Loes, Manon, Marjolein, Mirjam, Nienke P, Patricia, Peter, Susanne en Wouter kun

je je competing-risks dilemma nog bespreken? De post-WMM koffie maakt de woensdagochtend vervolgens compleet.

Het MMB-fietsclubje: wat plezier is het om na een dag hard nadenken met jullie het hoofd even leeg te rijden. De dagen worden alweer korter en ik hoop dat we volgend voorjaar weer vele na-het-werk-rondjes-Heuvelrug zullen rijden.

Alle andere AIOs, AIOS en andere collega's van de microbiologie, dank voor de gezelligheid op de afdeling. Van jullie heb ik geleerd mijn data niet als tabellen maar als figuren te presenteren en de *coffee-mailing-list* zorgt voor extra arbeidsvreugde.

Lieve Evelien B, lieve Nienke P, wat fijn dat jullie als paranimf naast mij willen staan tijdens de grote dag. Dank dat ik op jullie steun kan bouwen – nog even en dan staan jullie hier ook!

Lieve SUMMA-vriendinnen – Annelot, Marjolein, Fenna – na onze studie gingen wij inhoudelijk ieder onze eigen weg; toch blijven wij ervaringen delen en contact houden over kleine en grote afstanden. Zo'n vriendschap is veel waard.

Anne-Claar, Carolien, Jasmijn, Jetske, Linda en Sabine – lieve jaarclub – na 10 jaar is de maandagavond nog altijd gereserveerd om lief en leed met elkaar te delen. De weg naar dit boekje toe kennen jullie maar al te goed. Dank voor jullie onvoorwaardelijke steun tijdens de zowel de ups als downs! Op naar nog vele levensavonturen.

Lieve familie, jullie steun en interesse in wat ik doe zijn voor mij zo belangrijk. Lieve Elise, net als kunst en wetenschap, zijn wij soms zo verschillend en soms zo hetzelfde. Wat fijn dat je je creativiteit in mijn kaft hebt willen steken!

Lieve Jan en Reina, papa en mama, jullie hebben meer bijgedragen aan dit boekje dan jullie zelf waarschijnlijk denken. Altijd staan jullie voor mij klaar, en de vele kansen die jullie mij hebben gegeven hebben mij gebracht tot waar ik nu ben. Jullie hebben mij aangemoedigd om het onderste uit de kan te halen en uitdagingen aan te gaan. Bedankt!

Maaike

CURRICULUM VITAE

Maaike van Mourik was born on October 20th 1985 in Zeist, the Netherlands. She grew up partly in France and partly in the Netherlands. After completing her secondary school education (atheneum) at the Montessori Lyceum Herman Jordan in 2003, she enrolled at the University College Utrecht. During this time, she completed one semester of coursework at the University of Wisconsin-Madison and in 2006 she obtained her Bachelor of Science degree *summa cum laude*, majoring in biomedical sciences and completing a minor in psychology. She obtained her medical training in the Selective Utrecht Medical Master (SUMMA) at the University Medical Center Utrecht. As a medical student she had the opportunity to undertake an internship at the World Health Organization in Geneva. After her final clinical rotation in the department of internal medicine she started her scientific internship with the Department of Medical Microbiology on automated surveillance of healthcare-associated infections. She was awarded an Alexander Suermann research stipend and after completing her medical training in December 2010 (*cum laude*) she started the doctoral work presented in this thesis in January 2011 under supervision of prof.dr. M.J.M. Bonten, prof.dr. K.G.M. Moons and dr. A. Troelstra. As part of this project she spent several months as a visiting researcher with the Department of Population Medicine of Harvard Medical School and Harvard Pilgrim Healthcare in Boston under the supervision of dr. M. Klompas. She also completed the postgraduate Master in Clinical Epidemiology at Utrecht University. Besides her work as a PhD student, she serves on the board of the Young Alumni Network of Utrecht University. Maaike will continue as a postdoctoral researcher with the department of Medical Microbiology before pursuing further clinical training.

LIST OF PUBLICATIONS

Publications related to this thesis

Van Mourik MSM, Groenwold RHH, Berkelbach van der Sprenkel JW, Van Solinge WW, Troelstra A, Bonten MJM. Automated Detection of External Ventricular and Lumbar Drain-Related Meningitis Using Laboratory and Microbiology Results and Medication Data. *PLoS ONE 2011 6(8):e22846.*

Van Mourik MSM, Moons KGM, Van Solinge WW, Berkelbach van der Sprenkel JW, Regli L, Troelstra A, Bonten MJM. Automated detection of healthcare associated infections: External validation and updating of a model for surveillance of drain-related meningitis. *PLoS ONE 2012 7(12): e51509.*

Van Mourik MSM, Troelstra A, Van Solinge WW, Moons KGM, Bonten MJM. Automated surveillance for healthcare-associated infections: opportunities for improvement. *Clin Infect Dis 2013 Vol. 57(1): 85 - 93.*

Van Mourik MSM, Troelstra A, Berkelbach van der Sprenkel JW, Van der Jagt-Zwetsloot MCE, Nelson JH, Vos P, Arts MP, Dennesen PJW, Moons KGM, Bonten MJM. Validation of an automated surveillance approach for drain-related meningitis: a multicenter study. *Under revision.*

Van Mourik MSM*, Klein Klouwenberg PMC*, Ong DSY, Horn J, Schultz MJ, Cremer OL, Bonten MJM. Validation of a novel surveillance paradigm for ventilator-associated events. *Am J Resp Crit Care 2014 Vol. 189(8): 947 – 955.*

Van Mourik MSM, Moons KGM, Murphy MV, Bonten MJM, Klompas M. Case-mix adjustment for comparison of outcomes in mechanically ventilated patients using electronic routine care data. *Under revision.*

Van Mourik MSM, Troelstra A, Moons KGM, Bonten MJM. Accuracy of hospital discharge coding data for the surveillance of drain-related meningitis. *Infect Control Hosp Epidemiol 2013 Vol. 34(4): 433 – 436.*

Van Mourik MSM, Van Duijn PJ, Lee GM, Moons KGM, Bonten MJM. Accuracy of administrative data for surveillance of healthcare-associated infections: a systematic review. *Manuscript in preparation.*

Publications not related to this thesis

Van Mourik MSM, Macklon NS, Heijnen CJ 2009 Embryonic implantation: cytokines, adhesion molecules and immune cells in establishing an implantation environment. *J. Leukocyte biology 2009 Vol. 85: 4-19.*

Van Mourik MSM, Kamphuis MH, Plantinga NL, Croiset G, Ten Cate ThJ. Kennisniveau van studenten geneeskunde in een vierjarige en een zesjarige artsopleiding. *Tijdschrift voor Medisch Onderwijs 2010 Vol 29: 191-199.*

Van Mourik MSM, Cameron A, Ewen M, Laing RO. Availability, Price and Affordability of Cardiovascular Medicines: A comparison across 36 countries using WHO/HAI data. *BMC Cardiovascular disorders 2010 Vol. 10:25.*

Buelow E, Gonzalez TB, Versluis D, Oostdijk EA, Ogilvie LA, van Mourik MS, Oosterink E, van Passel MW, Smidt H, D'Andrea MM, de Been M, Jones BV, Willems RJ, Bonten MJ, van Schaik W. Effects of selective digestive decontamination (SDD) the gut resistome. *J Antimicrob Chemother 2014 In press.*