

EUR Research Information Portal

Impact of guidelines for the management of minor head injury on the utilization and diagnostic yield of CT over two decades, using natural language processing in a large dataset

Published in:
European Radiology

Publication status and date:
Published: 01/01/2019

DOI (link to publisher):
[10.1007/s00330-018-5954-5](https://doi.org/10.1007/s00330-018-5954-5)

Document Version
Publisher's PDF, also known as Version of record

Citation for the published version (APA):

Pons, E., Foks, K., Dippel, D., & Hunink, M. (2019). Impact of guidelines for the management of minor head injury on the utilization and diagnostic yield of CT over two decades, using natural language processing in a large dataset. *European Radiology*, 29(5), 2632-2640. <https://doi.org/10.1007/s00330-018-5954-5>

[Link to publication on the EUR Research Information Portal](#)

Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.



Impact of guidelines for the management of minor head injury on the utilization and diagnostic yield of CT over two decades, using natural language processing in a large dataset

Ewoud Pons^{1,2} · Kelly A. Foks^{3,4} · Diederik W. J. Dippel⁴ · M. G. Myriam Hunink^{1,5,6}

Received: 15 July 2018 / Revised: 2 November 2018 / Accepted: 4 December 2018 / Published online: 14 January 2019
© The Author(s) 2019

Abstract

Objectives We investigated the impact of clinical guidelines for the management of minor head injury on utilization and diagnostic yield of head CT over two decades.

Methods Retrospective before-after study using multiple electronic health record data sources. Natural language processing algorithms were developed to rapidly extract indication, Glasgow Coma Scale, and CT outcome from clinical records, creating two datasets: one based on all head injury CTs from 1997 to 2009 ($n = 9109$), for which diagnostic yield of intracranial traumatic findings was calculated. The second dataset (2009–2014) used both CT reports and clinical notes from the emergency department, enabling selection of minor head injury patients ($n = 4554$) and calculation of both CT utilization and diagnostic yield. Additionally, we tested for significant changes in utilization and yield after guideline implementation in 2011, using chi-square statistics and logistic regression.

Results The yield was initially nearly 60%, but in a decreasing trend dropped below 20% when CT became routinely used for head trauma. Between 2009 and 2014, of 4554 minor head injury patients overall, 85.4% underwent head CT. After guideline implementation in 2011, CT utilization significantly increased from 81.6 to 87.6% ($p = 7 \times 10^{-7}$), while yield significantly decreased from 12.2 to 9.6% ($p = 0.029$).

Conclusions The number of CTs performed for head trauma gradually increased over two decades, while the yield decreased. In 2011, despite implementation of a guideline aiming to improve selective use of CT in minor head injury, utilization significantly increased.

Key Points

- Over two decades, the number of head CTs performed for minor, moderate, and severe head injury gradually increased, while the diagnostic yield for intracranial findings showed a decreasing trend.
- Despite the implementation of a guideline in 2011, aiming to improve selective use of CT in minor head injury, utilization significantly increased, while diagnostic yield significantly decreased.
- Natural language processing is a valuable tool to monitor the utilization and diagnostic yield of imaging as a potential quality-of-care indicator.

Ewoud Pons and Kelly A. Foks contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-018-5954-5>) contains supplementary material, which is available to authorized users.

✉ Kelly A. Foks
k.foks@erasmusmc.nl

¹ Department of Radiology and Nuclear Medicine, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

² Department of Medical Informatics, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

³ Department of Public Health, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands

⁴ Department of Neurology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

⁵ Department of Epidemiology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

⁶ Center for Health Decision Sciences, Harvard T.H. Chan School of Public Health, Boston, USA

Keywords Craniocerebral trauma · Natural language processing · Data mining · Validation studies · Quality assurance, health care

Abbreviations

CCHR	Canadian CT Head Rule
CHIP	CT in head injury patients
ED	Emergency departments
EHR	Electronic health record
GCS	Glasgow Coma Scale
MHI	Minor head injury
NICE	National Institute for Health and Care Excellence
NLP	Natural language processing

Introduction

Non-contrast head CT is routinely used to rule out intracranial complications after (blunt) head trauma [1], but for patients with minor head injury (MHI) or mild traumatic brain injury—Glasgow Coma Scale (GCS) ≥ 13 —CT is not always necessary [2]. Intracranial traumatic findings are seen on 7–12% of CTs, although less than 1% of MHI patients require surgery, due to severe complications such as intracranial hematomas [3–6]. Over time, several guidelines have been developed to assess the risk of intracranial complications, using patient characteristics at presentation, such as vomiting or amnesia [3–5, 7]. These guidelines enable selective use of CT, with the goal to avoid unnecessary imaging and therefore reduce utilization. When comparing commonly used guidelines for MHI, the inherent trade-off between sensitivity and specificity with varying cutoff criteria is seen, leading to variation in the number of unnecessary head CTs and missed intracranial findings [8].

The purpose of implementing guidelines is to promote appropriate utilization which leads to safe, cost-effective practice that provides high-quality patient care. In the context of MHI, guidelines commonly reduce utilization. Nevertheless, several studies reported increased utilization of CT after guidelines for selective use were implemented [9, 10], leading to higher costs, longer waiting times, and additional radiation risk [11–13]. After implementation of validated imaging guidelines, it is important to assess their effectiveness in routine clinical practice. Both utilization (i.e., the proportion of patients that undergo imaging) and diagnostic yield (i.e., the proportion of imaging procedures with relevant findings) are important indicators for appropriate use of imaging.

The study purpose is to assess the impact of imaging guidelines for the management of MHI in routine clinical practice, by measuring both utilization and diagnostic yield of CT over two decades. We hypothesized that implementation of improved guidelines for selective use of CT would result in decreased utilization and consequentially also increased diagnostic yield over time.

The large number of clinical records related to MHI in this timeframe made manual review unfeasible. Natural language processing (NLP) can be used to extract structured variables from electronic free text and has been successfully applied to various sources in the electronic health record (EHR) [14], including radiology reports [15]. Therefore, NLP methods were developed to facilitate large dataset analytics of two decades of EHR sources.

Methods

We performed a retrospective before-after study using multiple EHR data sources from an urban, academic, level 1 trauma center for MHI patients presenting at the emergency department (ED). Part of the data was prospectively collected in the CT in Head Injury Patients (CHIP) study [4].

Sources and data collection

Several data sources related to MHI were obtained from the EHR, containing information on presentation, diagnostic imaging results, and other potentially relevant clinical outcomes: these sources included clinical notes from neurology, non-contrast head CT reports, neurosurgery registrations, hospitalization records, and various metadata (i.e., age, gender, and time of death for deceased patients).

NLP development and performance assessment

Four NLP algorithms were developed to:

1. Select acute head trauma cases from clinical notes;
2. Extract GCS score from clinical notes;
3. Select reports ordered for traumatic indication from all head CTs; and
4. Select head CT reports describing any intracranial traumatic finding.

Each NLP algorithm was trained on a set of reference documents, for which two or more clinicians manually labeled all information that should be extracted by NLP. The NLP algorithms for selecting acute head traumas and extracting GCS score were both trained using 500 labeled clinical notes from presentation. Additionally, traumatic indication was manually labeled in 500 head CTs, which were used for training the third NLP algorithm, in order to select traumatic cases from radiology reports directly—before clinical notes were documented electronically in time. Finally, 1934 CT head reports from 2002 to 2003 that had been labeled by our institute

during the CHIP study [4] were used to train the fourth NLP algorithm. Therefore, this algorithm selects CT reports with any intracranial traumatic finding (i.e., depressed fracture, subdural hematoma, epidural hematoma, subarachnoid hemorrhage, (non)hemorrhagic contusion, diffuse axonal injury, and intraventricular hemorrhage).

The first NLP algorithm for selecting acute head injuries was optimized for sensitivity to ensure completeness of the data. The fourth algorithm was optimized to balance false positives and false-negative detection of intracranial traumatic findings in a one-to-one ratio, to prevent potential changes in prevalence. During NLP development, 10-fold cross-validation was performed on the labeled reference sets, calculating sensitivity and specificity to measure the performance of all four NLP algorithms.

Dataset creation and validation

After performance evaluation, the four NLP algorithms were applied to all available clinical records (of the type used for training) to extract structured information. Radiology reports were available in digital format from 1997, while clinical notes only existed in the EHR from 2008. Therefore, the extracted variables were grouped into two distinct datasets (Fig. 1).

The first dataset was created by using the NLP algorithms three and four on all radiology reports from 1997 to 2009. This dataset contains minor, moderate, and severe head injury patients, containing CT reports as well as conventional X-ray of the head, which historically had been the first diagnostic test in the workup of head trauma.

The second dataset was created from both CT and clinical reports from 2009 to 2014, using NLP algorithms one, two, and four. Patients with GCS score < 13 were discarded, purposely resulting in a MHI dataset. This dataset also contained all clinical outcomes occurring within 30 days of presentation: hospitalization, neurosurgical intervention, and death. These outcomes were manually checked to ensure no critical lesions were missed by the initial head CT. Furthermore, integrity of this dataset was assessed by inspecting 100 randomly selected entries for completeness and correctness.

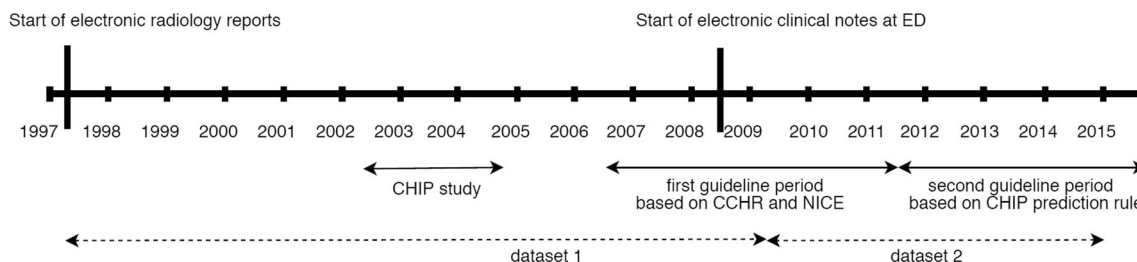


Fig. 1 Timeline of guidelines used in the study center and the generated datasets. Dataset 1 contains data extracted from electronic radiology reports between 1997 and 2009 for patients with minor, moderate, and severe head injury; dataset 2 contains data extracted from electronic

Guideline implementation over time

During the study timeframe, different diagnostic guidelines for MHI were used (Fig. 1). Until 2002, CT was mainly performed in MHI patients after detection of skull fractures on X-ray. From 2002 to 2004, the study center conducted the prospective CHIP study to investigate the risk factors of MHI, during which patients with GCS score of 13–14 and all patients with GCS score of 15 and at least one risk factor underwent CT (Appendix 1).

In 2006, the first local MHI guideline was implemented. This guideline was based on the Canadian CT Head Rule (CCHR) and the National Institute for Clinical Excellence (NICE) guideline to safely reduce CT utilization [3, 5]; CT was indicated in patients with GCS 15 and one risk factor, or a combination of specific risk factors (Appendix 1).

In 2011, the second MHI guideline was implemented based on the CHIP rule [4]. This guideline was developed to achieve a higher reduction in CTs, while identifying all patients with serious complications that require surgery; CT was indicated for patients with one major criterion or two minor criteria (Appendix 1).

The guideline implementation process remained stable over the years; at the study center, guidelines were based on national guidelines and developed in multidisciplinary groups, and regular updates were performed. The guidelines were presented to the involved clinical departments, formally approved by department staff and could easily be consulted online.

Statistical analysis

We calculated the diagnostic yield for both datasets by taking the proportion of positive findings from all CTs performed after trauma. Additionally, in the second dataset, we calculated utilization as the proportion of all MHI patients who underwent CT. The second dataset was split into two periods: period one, before implementation of the new CHIP-based guideline (June 2009–September 2011) and period two, after implementation (June 2012–September 2014). The datasets contained the same months to prevent bias due to seasonal variation. Furthermore, the datasets were separated by nine months to ensure the second

radiology reports and electronic clinical notes between 2009 and 2014, only for patients with minor head injury. CHIP, CT in head injury patients; CCHR, Canadian CT Head Rule; NICE, National Institute for Health and Care Excellence

guideline was fully operational at the start of the period. Descriptive statistics for patient demographics and outcomes were generated. We calculated the chi-squared statistic to compare both the utilization and yield of CT between the two periods. We performed logistic regression for the effect of time on both utilization and yield during each period independently, to test whether any significant trend existed within the periods. Finally, we compared the outcomes with the results of the CHIP study in the study center. Statistical analysis was performed with R software, version 3.3.2.

Results

Sources and data collection

We obtained 17,237 clinical notes documented by neurology in the ED, 27,759 non-contrast head CT reports, 2088 conventional skull X-ray reports, 10,207 neurosurgical procedure registrations, 4497 hospitalizations, and 2404 records of patients who had died (i.e., irrespective of the cause of death).

NLP performance assessment

NLP performance on 500 manually labeled clinical notes showed a 93.7% sensitivity and 97.4% specificity for the selection of acute head trauma cases, and a 97.5% sensitivity and 100% specificity for extraction of GCS score. Traumatic indication was determined with 95.8% sensitivity and 95.5% specificity on 500 manually labeled head CTs. Intracranial traumatic findings were identified with 86.8% sensitivity and 98.8% specificity on 1943 labeled head CT reports from the CHIP study. NLP errors during performance evaluation increased the tested prevalence by merely 0.25% compared to the training data.

Dataset creation and validation

The first dataset, based only on 18,606 radiology reports from 1997 to 2009, consisted of 9109 patients with a head CT for a traumatic indication. The second dataset, based on 9153 radiology reports and 17,237 clinical reports from 2009 to 2014, consisted of 4554 MHI patients.

After inspection of 100 patients in the second dataset, we found eight patients in which the NLP algorithms identified incorrect information from the clinical records. Three were incomplete due to extraction errors (a positive scan was missed once, while an incorrect GCS was selected twice). In one patient, imaging was scheduled according to the clinical notes, but the CT report was unavailable. NLP failed to exclude two trauma patients without apparent head injury and included one patient with a previous trauma in the history. One patient was incorrectly selected after transfer from another hospital. These results are consistent with the NLP

performance evaluation. Inspection of the follow-up outcomes within 30 days did not identify any misdiagnosed intracranial traumatic findings.

Dataset 1: Historical perspective of diagnostic yield for trauma of any severity (1997–2009)

Of 9109 patients who underwent a CT after sustaining a head injury, 18.0% ($n = 1641$) had intracranial traumatic findings on CT. Over time, more CTs were performed whereas the amount of skull X-rays diminished (Fig. 2a). During the early years, a low number of CTs were performed, most of which were positive resulting in a very high diagnostic yield. From 1997, the yield was initially nearly 60%, but a decreasing trend consolidated below 20% around 2002 (Fig. 2b), which illustrates that CT had become routinely used for head trauma. The effect of the CHIP study is somewhat noticeable in the lower yield associated with scanning all patients. In subsequent years, more CTs were performed with a relatively constant number of positive findings, resulting in a lower yield (Fig. 2a, b).

Dataset 2: Utilization and diagnostic yield of CT in MHI patients (2009–2014)

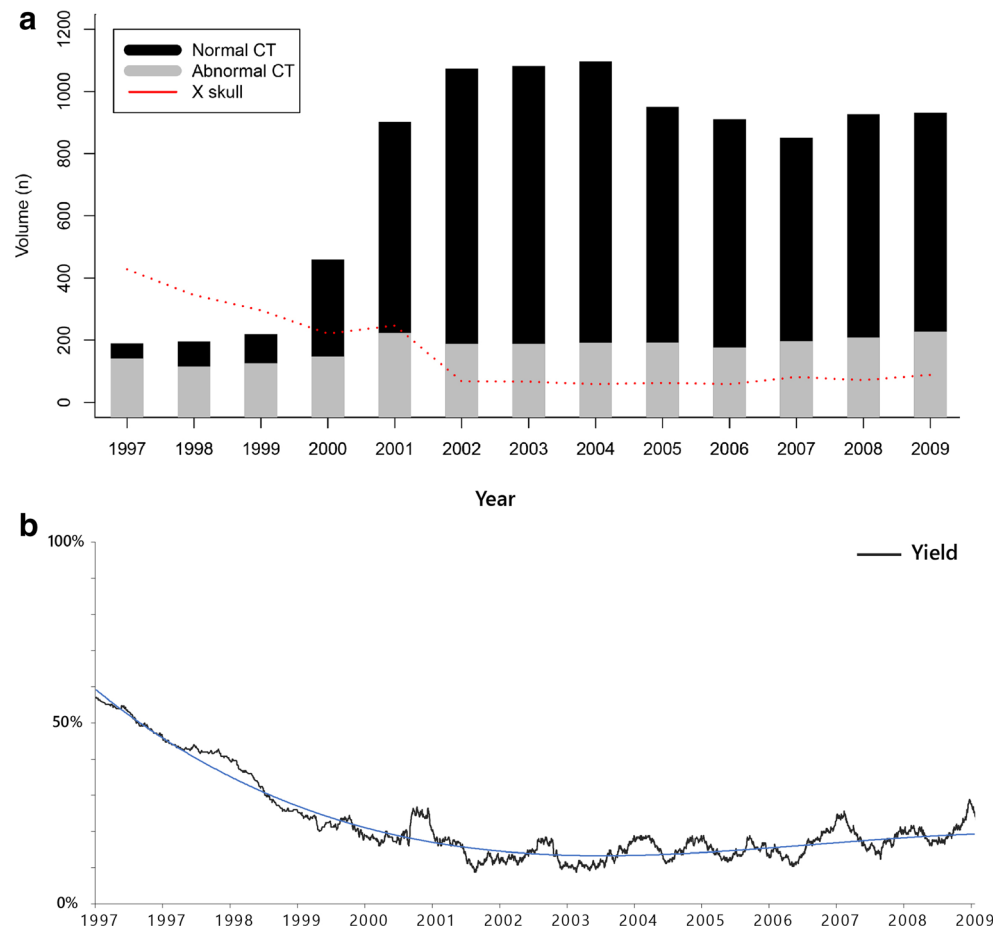
For 4554 patients with MHI seen at the ED, the mean age of 45.1 (SD \pm 20.3) years and most patients had GCS 15 ($n = 3219$; 70.7%) at presentation (Table 1). CT was performed in 3887 patients (85.4%), identifying 414 (9.1%) intracranial traumatic findings. Over time, the utilization of CTs in MHI increased, and the absolute number of positive findings on CT was stable, resulting in a decreasing diagnostic yield (Fig. 3a). Nine hundred seventy-seven patients (20%) were admitted to the hospital wards, and eight patients (0.18%) had a neurosurgical intervention within 30 days after injury. None of the patients without a head CT had a neurosurgical intervention within 30 days after the injury.

Impact of CT guidelines (second dataset)

During the CHIP study from 2002 to 2004, 2193 patients received a CT and 155 (7.1%) had intracranial traumatic findings (4). Eleven patients (0.50%) received a neurosurgical intervention within 30 days after injury.

The first guideline was used in 1429 patients from 2009 to 2011 (Table 1). Most patients were referred by ambulance ($n = 944$; 66.1%); only 26 patients were referred by a general practitioner (1.8%) and 370 patients (25.9%) came to the ED at their own initiative. In 1166 patients (81.6%), a head CT was performed and 142 (9.9%) had intracranial traumatic findings. The overall yield for the first period was 12.2%. Seven hundred thirteen patients (49.9%) were hospitalized and three of 1429 patients (0.21%) underwent a neurosurgical intervention within 30 days after injury.

Fig. 2 Historical perspective of CT use in patients with minor, moderate, and severe head injury from 1997 to 2009. **a** Number of patients with minor, moderate, and severe head injury and CT. The red line corresponds to skull X-ray performed for both traumatic and non-traumatic indications. **b** Yield of CT in minor, moderate, and severe head injury patients. To calculate yield for an exact point in time, we used a smoothed average of 125 entries before and after that date to calculate the proportion of positive findings



The second guideline was used in 2265 patients from 2012 to 2014. One thousand five hundred two patients (66.3%) were referred by ambulance, 38 (1.7%) by a general practitioner, and 614 (27.1%) came at their own initiative. In 1984 patients (87.6%) a CT was performed, 191 patients (8.4%) had intracranial traumatic findings. The overall yield for the second period was 9.6%. Nine hundred ninety patients (43.7%) were hospitalized, and three patients (0.13%) underwent a neurosurgical intervention within 30 days after injury.

The overall increase in utilization and decrease in diagnostic yield between the two guideline periods were both statistically significant (utilization $p = 7 \times 10^{-7}$ and yield $p = 0.029$). Within the periods individually, we found a slightly increasing trend of yield, and during the second period, a slightly decreasing trend of yield (Fig. 3b). Both slopes were not statistically significant compared to zero (Table 2).

Discussion

This study investigated the impact of clinical guidelines for the management of MHI in routine clinical practice, by assessing both utilization of CT and diagnostic yield for intracranial findings in all available electronic patient

records from two decades, facilitated by NLP. After implementation of a CHIP-based guideline in 2011, the utilization of CT increased significantly, while the yield significantly decreased. Within the periods before and after the guideline change, no significant trend was found for both utilization and yield, indicating that the before-after difference can be attributed to the guideline change and is not due to a preexisting trend. Therefore, implementation of improved guidelines for selective use of CT did not reduce utilization as we expected.

Our center conducted the CHIP study from 2002 to 2004, during which all MHI patients and at least one risk factor underwent CT [4]. When all the patients were scanned, the yield is approximately equal to the prevalence. This enabled us to compare the prevalence from the CHIP study period with the yield resulting from guideline use in routine clinical practice. The yield of CT during the CHIP study was 7.1%, whereas between 2009 and 2014, selective use of CT caused a higher yield of 10.7%. The overall prevalence of intracranial traumatic findings during both guideline periods was higher (9.1%) compared to the CHIP study (7.1%). The case mix during the CHIP study may have been slightly different from patients seen by neurology at the ED in routine clinical care. Importantly, this difference is not applicable to the guideline

Table 1 Characteristics of patients before and after implementation of a new minor head injury guideline

	CHIP study ^e (n = 2193)	First period (n = 1429)	Second period (n = 2265)	Entire cohort (n = 4554)
Period	February 2002–August 2004	June 2009–September 2011	June 2012–September 2014	June 2009–September 2014
Mean	1575 (71.8%)	1051 (73.5%)	1536 (67.8%)	3196 (70.2%)
Mean age in years (SD)	40.3 (± 18.1)	43.4 (± 20.1)	46.5 (± 20.5)	45.1 (± 20.3)
Emergency department				
GCS 13	106 (4.8%)	109 (7.6%)	116 (5.1%)	291 (6.4%)
GCS 14	387 (17.6%)	414 (29.0%)	440 (19.4%)	1044 (22.9%)
GCS 15	1661 (75.7%)	906 (63.4%)	1709 (75.5%)	3219 (70.7%)
Use of CT ^a	2193 (100%)	1166 (81.6%)	1984 (87.6%)	3887 (85.4%)
Any intracranial traumatic finding on CT (prevalence)	155 (7.1%)	142 (9.9%)	191 (8.4%)	414 (9.1%)
Yield of CT ^b	7.1%	12.2%	9.6%	10.7%
Follow-up				
Admission to hospital ^c	–	713 (49.9%)	990 (43.7%)	2067 (45.4%)
Neurosurgical intervention (< 30 days after injury)	11 (0.50%)	3 (0.21%)	3 (0.13%)	8 (0.18%)
Death (< 30 days after presentation) ^d	–	8 (0.56%)	21 (0.93%)	33 (0.72%)

^a Proportion of CT use in minor head injury patients. ^b Fraction positive findings. ^c Reason for admission to hospital unknown. ^d Unknown cause of death. ^e Patients with minor and at least one risk factor were included in this study. *CHIP*, CT in head injury patients; *GCS*, Glasgow Coma Scale; *ED*, emergency department; *CT*, computed tomography

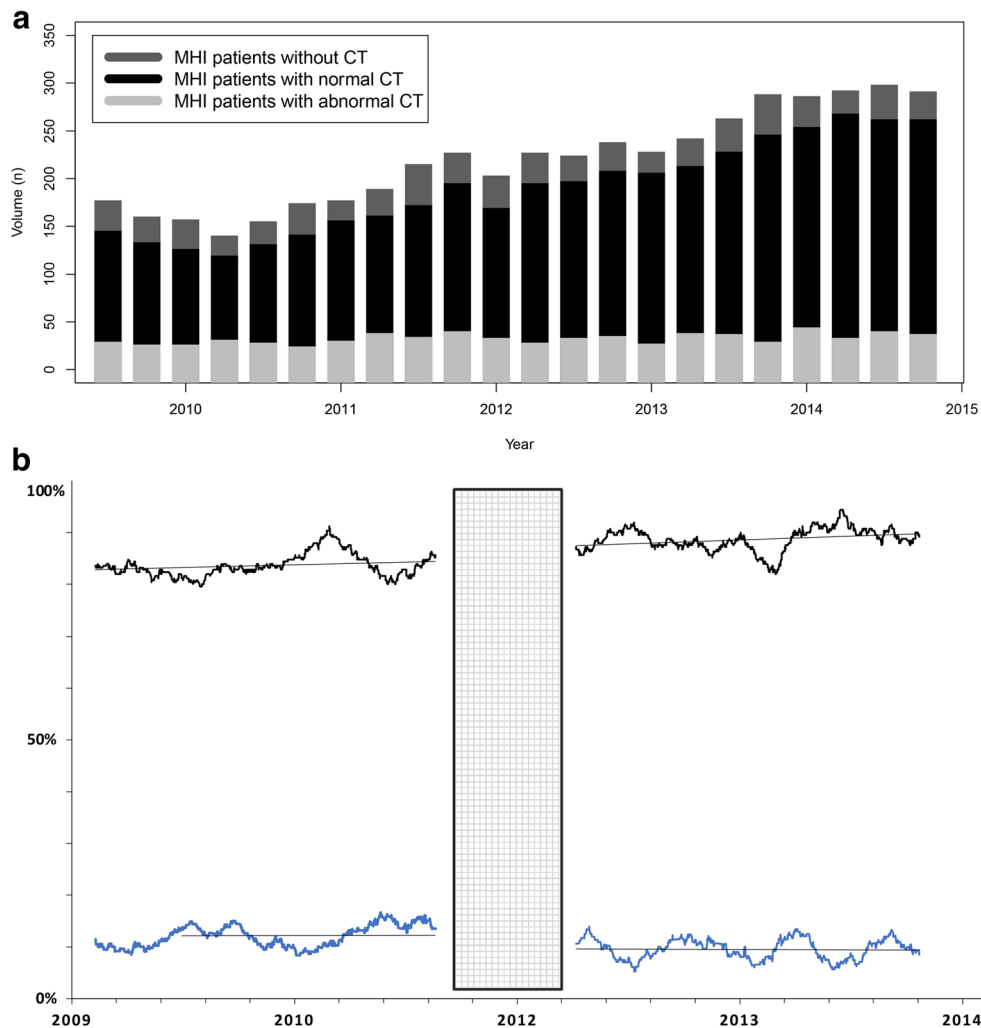
comparison, and both percentages are in line with the known incidence of intracranial traumatic findings [3–5].

The first guideline employed in our center reduced CT utilization by 18.4% compared to scanning all patients. In 2011, the CHIP-based guideline was implemented, and the potential CT reduction compared to scanning all patients was estimated at 23–30%, but we only showed a CT reduction of 12.4%. This lower reduction might be explained by more lenient use of the CHIP criteria in routine clinical practice. While during the first guideline period, the scanning policy was effectively stricter with significantly better CT reduction; not all patients with intracranial traumatic findings were identified [8, 16, 17]. Although the risk of serious traumatic findings requiring surgery is very low, scanning all MHI patients is more cost-effective than missing only a small portion of serious traumatic findings due to selective scanning [18, 19]. Thus, if the CHIP rule facilitates detection of all serious traumatic findings, while reducing CT use by 12%, this would be preferred to scanning all patients. We have shown the effect of using a CHIP-based guideline for selective scanning in routine clinical practice, and similar results were shown in a recent external validation study with a substantial reduction in CTs in clinical practice [20]. In the hypothetical situation that CT guidelines had not been implemented, in all likelihood, all patients with at least one risk factor would be scanned similar to the CHIP study period. This would almost completely eliminate any potential CT reduction.

To evaluate the purported impact of guidelines, information about guideline adherence by clinical physicians is necessary. However, for our study period, this information was not available. Furthermore, guideline adherence may affect CT utilization; however, we have no reason to assume that implemented guidelines were treated differently in one of the periods. Previous small studies about adherence of different CT guidelines showed an adherence in 51–100% of the patients [21–24]. Guideline adherence in our center cannot be expected to be 100% for the study data, which might have led to a lower CT reduction than expected. Additionally, introduction of a new guideline may also have resulted in enhanced awareness among clinicians for the risk factors in MHI, which may have caused increased utilization of CT [25]. The purpose of guideline implementation is to optimize clinical practice and care. Therefore, guideline implementation does not necessarily lead to a decrease in imaging utilization—in fact, it may lead to an increase in imaging if previously underutilized.

Besides guideline use, other factors such as increased presentation and different referral patterns by a general practitioner or ambulance can influence CT use. We found that during the second period, the number of patients seen in the ED had increased substantially (from 2265 to 1429). This increase cannot be explained by a difference in case mix because both demographic characteristics, as well as referral patterns by general practitioners and ambulance personnel remained the same. However, the increase is in line with previous research based on national

Fig. 3 Use and yield of CT in minor head injury patients from 2009 to 2014, before and after implementation of the second guideline. **a** Number of patients with minor head injury and CT. **b** Use and yield of CT in minor head injury patients; the gray area denotes the timeframe for implementation of the second guideline. To calculate yield for an exact point in time, we used a smoothed average of 125 entries before and after that date to calculate the proportion of positive findings



registries which identified more head injury patients presenting to the ED [11]. It is unclear whether this is caused by a potential increase in risky behavior, a higher tendency to seek urgent medical care. Other potential reasons may be an improvement of existing imaging technology, fear of litigation, or change of institutional culture, for example attitudes towards risk of missing diagnoses. Because of the increasing number of patients and the increase in the use of imaging [26, 27], efficient use of CT scanning is now more required than ever [8].

Despite, or maybe because of, scanning more patients in the second period, the number of patients admitted to the hospital was lower (49.9% vs 43.7%), which may be favorable to healthcare

costs [28]. Increased CT use can lead to increased confidence among physicians that no serious injury is present and thus allows discharge from the ED without admission to the hospital. Increased use may reflect cost-conscious changes in management.

The strength of NLP enabled the extraction of large numbers of clinical variables from heterogeneous EHR sources. In prior studies, NLP was used to assess diagnostic yield by extracting the imaging outcome from radiology reports [29, 30]. To our knowledge, this is the first study that investigated both the utilization and yield of diagnostic imaging by automatically extracting the indication as well as imaging outcome from free text, using multiple NLP algorithms on heterogeneous EHR

Table 2 The effect of time on use and yield of CT (2009–2014), estimated with logistic regression, where β is the increase in log odds ratio per day

	First period (June 2009–September 2011)			Second period (June 2012–September 2014)		
	Overall %	β	<i>p</i>	Overall %	β	<i>p</i>
Use	81.6	1.09×10^{-4}	0.693	87.6	1.91×10^{-4}	0.473
Yield	12.2	2.96×10^{-4}	0.330	9.6	1.87×10^{-4}	0.649

sources. Automatic information extraction enables the review of large numbers of textual documents but, equivalent to manual chart review, is not faultless. However, this has been shown to have limited impact, because extraction of traumatic cases and GCS was successfully optimized for very high specificity, resulting in mostly true cases in the final database. Also, during the performance evaluation, the prevalence of intracranial findings increased only by 0.25% due to errors. Any remaining errors can be assumed to affect all periods equally.

To conclude, in this large study using NLP, we showed that the number of head CTs performed for head injury gradually increased over two decades, while the diagnostic yield for intracranial traumatic findings demonstrated a decreasing trend. In 2011, despite implementation of an updated guideline aiming to improve selective use of CT for MHI, utilization significantly increased, while diagnostic yield significantly decreased. NLP is a valuable tool to monitor utilization and diagnostic yield of imaging as a potential quality-of-care indicator.

Funding This study was supported by the Open Technology Programme of Technology Foundation STW.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Prof. Dr. Hunink.

Conflict of interest Dr. Hunink receives Royalties from Cambridge University Press for a textbook on Medical Decision Making, reimbursement of expenses from the European Society of Radiology (ESR) for work on the ESR guidelines for imaging referrals, reimbursement of expenses from the European Institute for Biomedical Imaging Research (EIBIR) for membership of the Scientific Advisory Board, and research funding from the American Diabetes Association and the Netherlands Organization for Scientific Research (NWO). Dr. Dippel is the recipient of research grants from Dutch Heart Foundation, Dutch Brain Foundation, and unrestricted grants from AngioCare BV, Medtronic/Covidien/EV3®, MEDAC GmbH/LAMEPRO, Penumbra Inc., Stryker®, and Top Medical/Concentric, all unrelated to the research described in the present study.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was not required for this observational study.

Ethical approval Institutional Review Board approval was obtained for use of the retrospective data collections.

Study subjects or cohorts overlap Some study subjects from 2002 to 2003 have been previously reported in in article by Smits et al [4].

Methodology

- Retrospective
- Observational before-after study
- All data was generated at one institution

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Vos PE, Battistin L, Birbamer G et al (2002) EFNS guideline on mild traumatic brain injury: report of an EFNS task force. *Eur J Neurol* 9(3):207–219
2. Borg J, Holm L, Cassidy JD et al (2004) Diagnostic procedures in mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J Rehabil Med* (43 Suppl):61–75
3. Haydel MJ, Preston CA, Mills TJ, Luber S, Blaudeau E, DeBlieux PM (2000) Indications for computed tomography in patients with minor head injury. *N Engl J Med* 343(2):100–105
4. Smits M, Dippel DW, Steyerberg EW et al (2007) Predicting intracranial traumatic findings on computed tomography in patients with minor head injury: the CHIP prediction rule. *Ann Intern Med* 146(6):397–405
5. Stiell IG, Wells GA, Vandemheen K et al (2001) The Canadian CT Head Rule for patients with minor head injury. *Lancet* 357(9266):1391–1396
6. af Geijerstam JL, Britton M (2003) Mild head injury - mortality and complication rate: meta-analysis of findings in a systematic literature review. *Acta Neurochir (Wien)* 145(10):843–850
7. Pandor A, Goodacre S, Haman S et al (2011) Diagnostic management strategies for adults and children with minor head injury: a systematic review and an economic evaluation. *Health Technol Assess* 15(27):1–202
8. Smits M, Dippel DW, de Haan GG et al (2007) Minor head injury: guidelines for the use of CT—a multicenter validation study. *Radiology* 245(3):831–838
9. van den Brand CL, Rambach AH, Postma R et al (2014) Practice guideline 'Management of patients with mild traumatic head/brain injury' in the Netherlands. *Ned Tijdschr Geneeskd* 158:A6973
10. Shrivat BP, Huseyin TS, Hynes KA (2006) NICE guideline for the management of head injury: an audit demonstrating its impact on a district general hospital, with a cost analysis for England and Wales. *Emerg Med J* 23(2):109–113
11. Van den Brand CL, Karger LB, Nijman STM, Hunink MGM, Patka P, Jellema K (2017) Traumatic brain injury in the Netherlands, trends in emergency department visits, hospitalization and mortality between 1998 and 2012. *Eur J Emerg Med*
12. Maas AIR, Menon DK, Adelson PD et al (2017) Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol* 16(12):987–1048
13. Brenner DJ, Hall EJ (2007) Computed tomography—an increasing source of radiation exposure. *N Engl J Med* 357(22):2277–2284
14. Kreimeyer K, Foster M, Pandey A et al (2017) Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform* 73:14–29
15. Pons E, Braun LM, Hunink MG, Kors JA (2016) Natural language processing in radiology: a systematic review. *Radiology* 279(2):329–343

16. Smits M, Dippel DW, de Haan GG et al (2005) External validation of the Canadian CT Head Rule and the New Orleans Criteria for CT scanning in patients with minor head injury. *JAMA* 294(12):1519–1525
17. Bouida W, Marghli S, Souissi S et al (2013) Prediction value of the Canadian CT head rule and the New Orleans criteria for positive head CT scan and acute neurosurgical procedures in minor head trauma: a multicenter external validation study. *Ann Emerg Med* 61(5):521–527
18. Holmes MW, Goodacre S, Stevenson MD, Pandor A, Pickering A (2012) The cost-effectiveness of diagnostic management strategies for adults with minor head injury. *Injury* 43(9):1423–1431
19. Smits M, Dippel DW, Nederkoorn PJ et al (2010) Minor head injury: CT-based strategies for management—a cost-effectiveness analysis. *Radiology* 254(2):532–540
20. Foks KA, van den Brand CL, Lingsma HF et al (2018) External validation of computed tomography decision rules for minor head injury: prospective, multicentre cohort study in the Netherlands. *BMJ* 362:k3527
21. National Clinical Guidance Centre (UK) (2014) Head injury: triage, assessment, investigation and early management of head injury in children, young people and adults. National Institute for Health and Care Excellence (UK)
22. Haydon NB (2013) Head injury: audit of a clinical guideline to justify head CT. *J Med Imaging Radiat Oncol* 57(2):161–168
23. Mooney JS, Yates A, Sellar L et al (2011) Emergency head injury imaging: implementing NICE 2007 in a tertiary neurosciences centre and a busy district general hospital. *Emerg Med J* 28(9):778–782
24. Heskestad B, Baardsen R, Helseth E, Ingebrigtsen T (2008) Guideline compliance in management of minimal, mild, and moderate head injury: high frequency of noncompliance among individual physicians despite strong guideline support from clinical leaders. *J Trauma* 65(6):1309–1313
25. Rohacek M, Albrecht M, Kleim B, Zimmermann H, Exadaktylos A (2012) Reasons for ordering computed tomography scans of the head in patients with minor brain injury. *Injury* 43(9):1415–1418
26. Bellolio MF, Heien HC, Sanglaralingham LR et al (2017) Increased computed tomography utilization in the emergency department and its association with hospital admission. *West J Emerg Med* 18(5): 835–845
27. Brinkjicki W, Kallmess DF, Cloft HJ (2015) Rising utilization of CT in adult fall patients. *AJR Am J Roentgenol* 204(3):558–562
28. Af Geijerstam JL, Britton M, Marké LA (2004) Mild head injury: observation or computed tomography? Economic aspects by literature review and decision analysis. *Emerg Med J* 21(1):54–58
29. Dreyer KJ, Kalra MK, Maher MM et al (2005) Application of recently developed computer algorithm for automatic classification of unstructured radiology reports: validation study. *Radiology* 234(2):323–329
30. Raja AS, Ip IK, Prevedello LM et al (2012) Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department. *Radiology* 262(2): 468–474