Title

Building Consensus for a Shared Definition of Adverse Events: A Case Study in the Profession of Dentistry

Journal

Journal of Patient Safety, 18(5)

ISSN

1549-8417

Authors

Franklin, Amy
Kalenderian, Elsbeth
Hebballi, Nutan
et al.

Publication Date

2022-08-01

DOI

10.1097/pts.0000000000000959

Peer reviewed

HHS Public Access

Author manuscript

J Patient Saf. Author manuscript; available in PMC 2023 August 01.

Published in final edited form as:

J Patient Saf. 2022 August 01; 18(5): 470–474. doi:10.1097/PTS.0000000000000959.

Building consensus for a shared definition of adverse events. A case study in the profession of dentistry

Amy Franklin[1], **Elsbeth Kalenderian**[2], **Nutan Hebballi**[3], **Veronique Delattre**[3], **Jini Etoule**[4], **Joel White**[2], **Ram Vaderhobli**[2], **Denice Stewart**[5], **Karla Kent**[6], **Alfa Yansane**[4], **Muhammad Walji**[3]

[1]School of Biomedical Informatics

[2]Department of Preventive and Restorative Dental Sciences, School of Dentistry, University of California San Francisco, San Francisco, CA, USA

[3]School of Dentistry, The University of Texas Health Science Center at Houston, Houston, TX, USA

[4]Oral Health Policy and Epidemiology, School of Dental Medicine, Harvard University, Boston, MA, USA

[5]Community Dentistry

[6]Integrative Biosciences, School of Dentistry, Oregon Health and Science University, Portland, OR, USA

Abstract

Background: To achieve high quality health care, adverse events (AEs) must be proactively recognized and mitigated. However, there is often ambiguity in applying guidelines and definitions. We describe the iterative calibration process needed to achieve a shared definition of AEs in dentistry. Our alignment process includes both independent and consensus building approaches.

Objective: We explore the process of defining dental AEs and the steps necessary to achieve alignment across different care providers.

Methods: Teams from four dental institutions across the United States iteratively reviewed patient records following identification of charts using an automated trigger tool. Calibration across teams was supported through negotiated definition of AEs and standardization of evidence provided in review. Inter-rater relability was assessed using descriptive and kappa statistics.

Results: Following five iterative cycles of calibration, the teams (n=8 raters) identified 118 cases. The average percent agreement for AE determination was 82.2%. Further, the average, pairwise prevalence and bias adjusted kappa (PABAK) was 57.5% (κ=0.575) for determining AE presence. The average percent agreement for categorization of the AE type was 78.5% while the PABAK

Corresponding Authors: Amy Franklin, PhD, The University of Texas Health Science Center at Houston, School of Biomedical Informatics, 7000 Fannin Suite 165, Houston, TX 77030, 713-486-3670 office, 713-500-3907 fax; Elsbeth Kalenderian, DDS, MPH, PhD, School of Dentistry, University of California San Francisco, 707 Parnassus Ave, San Francisco CA 94143.

was 48.8%. Lastly, the average percent agreement for categorization of AE severity was 82.2% and the corresponding PABAK was 71.7%.

Conclusions: Successful calibration across reviewers is possible following consensus building procedures. Higher levels of agreement were found when categorizing severity (of identified events) rather than the events themselves. Our results demonstrate the need for collaborative procedures as well as training in for the identification and severity rating of AEs.

Introduction

The study of adverse events (AEs) in healthcare has a significant body of literature exploring topics including prevalence, difficulty with AE reporting, and the effectiveness of methods for identifying events such as manual chart review,(1–3) and data mining of billing or malpractice claims data(4). Additionally, a developing body of work is exploring electronic health record mining(5, 6) along with automated tools for detection. Similar lines of research for studying adverse events in dentistry(7, 8) are emerging including creation of trigger tools to automatically identify the precursors of AEs(9) and descriptions of case studies(10) and reports of device failures(11). An important step in understanding and managing AEs is consistent identification of the adverse event. Here, we detail the process used to develop and achieve reliable use of a definition for AEs in dentistry across a distributed group of practitioners.

Defining Adverse Events (AEs)

Death, life-threatening events, and permanent damage are readily identified as adverse events. However, in dentistry, as in other areas of healthcare, there are other outcomes that exist in the gray(er) space between issues of quality, desirability of clinical outcomes and observable harm. Making the connections between an AE from an error, a near miss, or poor outcome of a treatment may be challenging. For example, there may be a fine line between an AE and a Quality of Care (QoC) issue based on the intention of clinicians and adherence to protocols. Additionally, less desirable outcomes following the natural course of the disease may be confused with adverse events (i.e. loss of sight with chronic diabetes, erosion of teeth in patients with bulimia.) To define what exactly is an adverse event, we turn to existing definitions and classifications from other areas of healthcare to inform our efforts in dentistry.

Looking across organizations such as the Agency for Healthcare Research and Quality (AHRQ), the Food and Drug Administration (FDA), and the Institute for Healthcare Improvement (IHI), we find variations in how adverse events are defined. Some agencies, such as AHRQ, define AEs by outcomes involving ***injury*** *caused by medical care*. Compare this to the Food and Drug Administration's view of adverse events as *any* ***undesirable experience*** *associated with the use of a medical product in a patient (12).* The potential space between injury and negative experience covers a gulf of quality issues and subjective opinion. While adverse events do not imply "error," "negligence," or "poor quality of care", it is unclear what other types of events might exist in the larger space created by the FDA definition. The Institute for Healthcare Improvement (IHI) provides greater specificity by requiring that an AE is *unintended* ***physical injury*** *resulting from or contributed to by*

medical care ***requires additional monitoring, treatment or hospitalization, or that results in death (23).*** This better defines the boundaries of the degree of injury and treatment required to achieve the classification of an AE.

Aspects of each of these definitions come into play as we move to determine what are the necessary and sufficient features of an adverse event in dentistry. The definition must be able to capture events from diagnostic error, harm resulting from a failure to comply or adhere to the standard of care, and to events such device failures.

Challenges to Identifying Adverse Events

The challenges of identifying adverse events are three fold. The first challenge is defining the scope of the adverse event, the next is recognizing the potential conditions for an AE to occur, and then the final hurdle is the classification and/or categorization of each event considering other similar occurrences. In short in order to say you have identified an AE, you have to be able to reliably apply a definition, across real cases, and with different people arriving to the same conclusion. It is the process of building a consensus of defining an event, determining what evidence or data is required to establish that case, and articulating a shared understanding that we outline here.

Sharing a Mental Model of Adverse Events in Dentistry

Building consensus and reliability requires a shared mental model. This sharing can be as minimal as an overlapping mental representation by members of a group to a more team-based or collaborative notion of mental models in work. In sharing, there is a shift from an individual knowledge structure (a la Johnson Laird(13)) to group level cognition(14). Just as Johnson-Laird(15) points out “models develop as an individual progresses from novice to expert” so through interaction there is modification and reinforcement of the collective model(16). It is through this application to real instances and later the discourse process of negotiating shared understanding of those events that our definition of adverse events emerges.

Within healthcare as a whole, there are conflicting results regarding how reliably definitions of adverse events can be applied. Although the degree varies by study and method, reliability across individuals in identifying AEs is typically only low to fair(3). Tools such as the global trigger tool (GTT)(2, 17) have been developed to support rapid identification and the measurement of these occurrences over time. Although the GTT has been shown to be more successful in identifying AEs when compared to self-reports or random chart review, the consistency of its application across clinicians or implementation sites is unclear. Schildmeijer et al (18) found in a comparison of five hospital systems the inter-rater reliability on similar chart reviews to range from slight (kappa=.26) to substantial (kappa=.77.) When interviewing users of the tool, Schildmeijer et al remarked “*small, gradual methodological changes together with continuingly developed expertise and adaptation to looking at harm from a patient’s perspective may contribute to large differences in assessment over time(19).*” These findings indicate how problematic it may be to get independent reviewers to draw similar conclusions regarding AEs. This challenge is

further strengthened by Mattsson et al who found that reviewers evaluating *the same charts* achieve only moderate (kappa =.45) inter-rater reliability(20).

When it comes to group assessment, two reviewers are not more effective than singletons(21). Consensus-based processes have shown higher reliability with group-generated responses as compared to a gold standard (22). This project employed a consensus-based process, with a hypothesis that rigorous discussion would lead to a shared mental model among the participants and thus greater calibration and consistency.

Methods

Setting

Four academic clinical sites within the US were involved in this study. At these sites patients can obtain dental care from predoctoral dental students, advanced graduate dental residents in the teaching practices (TPs) or from academic faculty members in the faculty practice (FP). All sites utilized the same Electronic Health Record (EHR) (axiUm, Exan, Coquitlam, British Columbia, Canada) which was queried as part of the record review. Each team consisted of two members who were trained dentists or had substantial experience in assessing dental quality. As part of the project, the research team is working together to develop a patient safety system for dentistry. This includes developing AE definitions, creating tools to identify and capture AEs, building a classification scheme to organize the AEs, and generating tools to enable the systematic analysis of AEs.

Triggering charts—Following a method akin to the IHI global trigger tool(23), automated methods were used to rapidly sample patient data by identifying “triggers” that may signal adverse events across modules of care. A “trigger” is an opportunity or clue to identify adverse events in patient’s dental record. For example, a second appointment on the same tooth within a short duration or prescription of an antibiotic days after a procedure may indicate a potential adverse event. However, the presence of triggers does not represent adverse events. Triggered charts, rather than random review of charts, were used to reduce the search space for dental adverse events.

We employed a method of using a trigger tool that selects dental cases based on the presence of certain standardized Current Dental Terminology (CDT)(24) codes, date ranges for visits and other details to such as the generation of prescriptions(9). SNODDS, previously called DDS (Dental Diagnostic System) as well as the EZCodes)(25–27) is the dental diagnostic terminology in place at the study sites. Triggers were then applied using specific rules for query in potential events. For example, a query might look for a crown, root canal therapy or restoration identified through the Current Dental Terminology (CDT) codes along with a CDT code for an extraction on the same tooth within 365 days. Following the automatic identification of potential AEs within triggered charts, clinical teams were tasked with the review of patient charts using a standardize form to determine the presence or absence of an adverse event considering the selected cases.

The review process

Five iterations of chart review were completed. In each round, triggered charts were reviewed using a standardized data collection tool. Each chart was evaluated by independent reviewers and screened for the presence of the appropriate trigger. Next, the presence or absence of an AE was noted. If an AE was identified each reviewer then documented the event by using a standardized data capture form. Documentation required summarizing the event, selecting the classification of the AE according to a predefined set of categories, and providing a rating regarding the severity of the AE.

The severity score, described in previous work (28), adapts the classification from the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index for Categorizing Errors(29) and is based on the IHI global trigger tool modifications. This taxonomy can be used to describe events ranging from no harm to harm resulting in death along a continuum. Examples of these categories include required monitoring to confirm no harm to the patient and/or required intervention to preclude harm (Category D),temporary harm requiring intervention (Category E), and permanent harm (Category G).

Following individual chart review, comparisons across reviewers were conducted locally within a team and in later stages across sites. In each iteration of review, group discussions including a case-by-case review of charts were used to support (re)design of operational definitions and modification to the review process. Reliability across reviewers is analyzed in the last three stages.

Consensus Procedures

Here, we compare the reliability of reviewers using kappa statistics as an assessment of the degree of shared understanding regarding the definition of AEs in dentistry. We describe the process of review as it changed over time from local to more group comparisons. We detail the shifting definition of adverse events and the modifications to both the chart review tool and evidence pulled from the patient record used in decision making. These changes are gathered from the shared weekly discourse on the chart review process. Weekly group discussions regarding chart review were conducted lasting from 90–120 minutes per meeting. These discussions were in addition to the collaborative interactions of reviewers within the same site as they went about completing their reviews of charts. While this method of assessment of reliability is a somewhat atypical means of capturing the overlap in reviewers' models of AEs, we believe changes in the language and evidence used by the group demonstrate the evolution of their shared model overtime. This overlap is reflected in improved reliability scores.

Results

Round 1

In the first round, adverse events were defined as ***harm*** *caused to the patient by dental care, regardless of whether it is associated with an error or is considered preventable.* Each site was asked to identify AEs within their own records (using different triggers) achieving

consensus between reviewers at each location. Two triggers, one on failed implant and a random trigger, were deployed across all sites. All other triggers had at least two schools complete the review for those specific types of cases. The positive predictive value for each trigger was calculated.[15] Site-specific comparison for reliability (i.e. overlap within reviewers at each location) was not calculated, as reviewers would compare chart reviews and reach consensus before sending results to the coordinating center.

When the results are compared for the success of specific triggers and the evidence used to determine the appropriateness of an AE designation, significant variation was discovered across and within sites. For example, a need for multiple visits regarding the same tooth varied in its positive predictive value. As comparisons were made using the same trigger with different charts and different reviewers (i.e. Trigger 6 at facility 1 and facility 2) it became clear that it was difficult for reviewers to consistently apply this definition in similar cases. Following group discussion, the chart review tool was modified to better fit the expectations of users regarding how to document the AE and the work process of shared review was modified. Additionally, following group collaboration regarding the identification of AEs, further refinement of the definition of what is considered an AE was required.

Round 2

In the second iteration of the manual review process, a modified definition of AEs and revised forms were deployed. The revised definition added greater specification that *injury resulting from a medical/dental intervention could not be* ***due to the underlying clinical condition of the patient.*** The continued focus on *harm* and *severity* remain signification conditions for the recognition of events. The changes in form and specification lead to modifications in the data being included in the group discussion and provided additional fodder for the consensus building process.

Round 3

In the third iteration of the alignment process, reviewers across sites used similar but not identical data. For example, while the reviewers were all asked to review similar type of events, i.e. post surgical complications involving infection, they did not review the same instances. Group discussion following review indicated differences in the application of the definition across different cases with variation at the site level. Little constraint was placed on the characterization of an event as adverse. Harm was the main criteria without specifying the degree of impact.

As group discussion followed individual and site efforts, akin to a modified Delphi procedure, a new classification system for AEs emerged. Now an adverse event was understood as, ***physical harm that is moderate or severe due to treatment within a timeframe relevant to the clinical scenario.*** Additionally and importantly, a now constrictive model emerged from the group regarding the definition. AEs were no longer solely defined by the presence of necessary and sufficient conditions (i.e. harm of certain extent within a given duration) but were also now defined by exclusion (Table 1 below). The new criteria limit the scope of AE to no longer include quality concerns in the absence

of harm. Specific instances of pain (e.g. slight), loss of function (that has no accompanied harm) or repeated treatments became named exclusions.

Following this cycle of calibration and redefinition, individual reviewers were asked to complete a re-review to determine the appropriateness of the new characterization of AEs to those occasions. Pairwise inter-rater reliability was completed and Light's kappa was determined to be 0.321 (fair.) In an effort to improve reliability further, an additional round of reliability was calculated.

Fourth Iteration

In the fourth iteration, limited set of reviewers (4) considered a subset of 27 charts (Table 2 below). The pairwise reliability dropped from the previous fair to now slight reliability at .0126.

A drop in reliability was unexpected at this point as the definition of AEs had become stricter. Through continued discussion of cases and maintenance of the definition including exclusion holding steady, it was determined that rather than modifying the definition further, at this point, there should be a shift in calibration techniques.

Reviewers were encouraged to achieve greater alignment through the review of materials. Selected ***shared*** cases were given to the entire group forcing a consistency across users in the data considered as part of their decision making process. Previously different assessment at each site using site-specific data may have included variation in the content of the patient record. The availability of information for mining may have been a contributing factor to the variation in applying the protocol for identifying AE. Using shared cases, standardizes the evidence available to the reviewer for decision-making. This shift was expected to improve reliability. Additionally, although not related to the definition of AEs, the trigger logic itself was also updated to better reflect the types of AEs shown to be uncovered with the prior coding of events.

Fifth Iteration

Following five iterative cycles of calibration, the four teams (n=8 raters) identified and reviewed an additional 118 cases. The average percent agreement for AE determination was 82.2%. Further, the average, pairwise prevalence and bias adjusted kappa (PABAK) was 57.5% (κ=0.575) for determining AE presence. The average percent agreement for categorization of the AE type 78.5% while the PABAK was 48.8%. Lastly, the average percent agreement for categorization of AE severity was 82.2% and the corresponding PABAK was 71.7%.

As evidenced by the discussion, not only did this new process improve overall reliability, it led to further constriction in the shared understanding of category specific information. Now seven of the twelve classifications for AEs have increased granularity in their definition. For example, pain as a category now has inclusion/exclusion criteria based on medication needs, severity of pain, and the presence of a dry socket. This additional level of information narrows the application of this category and allows for greater agreement across reviewers.

Discussion

Our results replicated and expanded on the medical literature demonstrating the challenges of achieving consistent categorization of adverse events. Like the previous work, we see evidence of an impact of experience and expertise in the reviewers, drift in scope of definitions, and differences in the application of definitions across sites. We demonstrate through the evolution of our processes that consistency across reviewers, sites, and instances requires careful and explicit management of the definitions, evidence and tools used in these efforts.

Outcomes from our work include modifications to our classification schema and training procedures. In addition to the u nfolding definition of adverse events broadly applied, the specific categories in which an AE might fall also transformed over time. The original classification of AEs used in this project[26] was generated through literature review, focus group interviews, and the insight of the authors. It included 19 originial categories such as toxicity, failure or malfunction of a device, procedure on the wrong patient, and nerve damage or harm. While initially this classification grew to include other types of harm such as prolonged bleeding and ocular damage, like the definition of AEs, ultimately the classification system was reduced to a set of twelve types seen in the table 1. Unlike the progression in definition that grew through exclusion and specificity, this schema was reduced as classifications were collapsed to be more encompassing (e.g. grouping wrong patient, wrong site, wrong procedure.)

Given the challenges in bringing a distributed group of practitioners into alignment regarding the definition of adverse events in dentistry, how to scale to other new reviewers and bigger systems is a challenge. As part of our group’s efforts we have created a training system for future reviewers that attempts to replicate the interaction of an individual with an “other” to challenge their use of the term AE and to question their understanding of necessary evidence in categorizing an event. Using a learning management systems (LMS), we have generated a set of protocols and training materials to guide a novice through the stages we have outlined here. First, they are trained on the terms and definitions provided. They are then asked to apply these ideas to real world cases. As with our consensus-building interaction, when the novice returns with their annotations in place to the LMS, they are met with potentially disparate opinions provided by the system. They are asked to provide for given cases identification of the triggers and adverse events, severity scores, demographic data and other components of the review. Using the quiz features of the learning system, we provide specific feedback to their responses attempting to shape their understanding to better fit the now group-defined gold standard. We are, through more independent work, attempting to replicate the discussions required for strong and consistent application of AE classifications to real events.

Conclusions

The shared understanding of what is an adverse event in dentistry morphed over time and through experience of the reviewers. The evolution was influenced by testing of the necessary and sufficient features through application on real cases. It was restructured

through group consensus building processes and required common evidence to achieve a reliable meaning. We demonstrated the progression of a formalization of the concept, here adverse events, to the reality of recognizing this effort in practice requires a team, time, and a great deal of discussion to achieve a consistent result.

Acknowledgements:

Research reported in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under Award Number R01DE022628. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med 1991;324(6):370–6. [PubMed: 1987460]
2. Ashton CM, Kuykendall DH, Johnson ML, Wray NP. An empirical assessment of the validity of explicit and implicit process-of-care criteria for quality assessment. Med Care 1999;37(8):798–808. [PubMed: 10448722]
3. Localio AR, Weaver SL, Landis JR, Lawthers AG, Brenhan TA, Hebert L, et al. Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. Ann Intern Med 1996;125(6):457–64. [PubMed: 8779457]
4. Phillips RL Jr., Bartholomew LA, Dovey SM, Fryer GE Jr., Miyoshi TJ, Green LA. Learning from malpractice claims about negligent, adverse events in primary care in the United States. Qual Saf Health Care 2004;13(2):121–6. [PubMed: 15069219]
5. Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. JAMA 1991;266(20):2847–51. [PubMed: 1942452]
6. Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. J Am Med Inform Assoc 1998;5(3):305–14. [PubMed: 9609500]
7. Ramoni RB, Walji MF, White J, Stewart D, Vaderhobli R, Simmons D, et al. From good to better: toward a patient safety initiative in dentistry. J Am Dent Assoc 2012;143(9):956–60. [PubMed: 22942131]
8. Ramoni R, Walji MF, Tavares A, White J, Tokede O, Vaderhobli R, et al. Open wide: looking into the safety culture of dental school clinics. J Dent Educ 2014;78(5):745–56. [PubMed: 24789834]
9. Kalenderian E, Walji MF, Tavares A, Ramoni RB. An adverse event trigger tool in dentistry: a new methodology for measuring harm in the dental office. J Am Dent Assoc 2013;144(7):808–14. [PubMed: 23813262]
10. Obadan EM, Ramoni RB, Kalenderian E. Lessons learned from dental patient safety case reports. J Am Dent Assoc 2015;146(5):318–26 e2. [PubMed: 25925524]
11. Hebballi NB, Ramoni R, Kalenderian E, Delattre VF, Stewart DC, Kent K, et al. The dangers of dental devices as reported in the Food and Drug Administration Manufacturer and User Facility Device Experience Database. J Am Dent Assoc 2015;146(2):102–10. [PubMed: 25637208]
12 . Available from: http://www.fda.gov/Safety/MedWatch/HowToReport/ucm053087.htm.
13. Johnson-Laird PN. Towards a Cognitive Science of Language, Inference, and Consciousness Cambridge, MA: Harvard University Press; 1983.
14. Klimoski R, S M. Team mental model: Construct or metaphor? Journal of management 1994;20(2):403–37.
15. Johnson-Laird PN. Mental Models Posner MI, editor. Cambridge, MA: MIT Press; 1989.
16. Mohammed S, Dumville BC. Mohammed S, & Dumville BC (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. Journal of organizational Behavior 2001;22(2):89–106.

17. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood) 2011;30(4):581–9. [PubMed: 21471476]
18. Schildmeijer K, Nilsson L, Arestedt K, Perk J. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. BMJ Qual Saf 2012;21(4):307–14.
19. Schildmeijer K, Nilsson L, Perk J, Arestedt K, Nilsson G. Strengths and weaknesses of working with the Global Trigger Tool method for retrospective record review: focus group interviews with team members. BMJ Open 2013;3(9):e003131.
20. Mattsson TO, Knudsen JL, Lauritsen J, Brixen K, Herrstedt J. Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. BMJ Qual Saf 2013;22(7):571–9.
21. Zegers M, de Bruijne MC, Wagner C, Groenewegen PP, van der Wal G, de Vet HC. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. J Clin Epidemiol 2010;63(1):94–102. [PubMed: 19473812]
22. Patterson PD, Lave JR, Weaver MD, Guyette FX, Arnold RM, Martin-Gill C, et al. A comparative assessment of adverse event classification in the out-of-hospital setting. Prehosp Emerg Care 2014;18(4):495–504. [PubMed: 24878451]
23. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series white paper [Internet] 2009 4/10/2015. Available from: http://www.IHI.org.
24. American Dental Association. CDT: Code on Dental Procedures and Nomenclature Chicago, Ill.2015 [Available from: http://www.ada.org/en/publications/cdt/.
25. Kalenderian E, Ramoni RL, White JM, Schoonheim-Klein ME, Stark PC, Kimmes NS, et al. The development of a dental diagnostic terminology. Journal of Dental Education 2011;75(1):68–76. [PubMed: 21205730]
26. Ramoni RB, Walji MF, Kim S, Tokede O, McClellan L, Simmons K, et al. Attitudes toward and beliefs about the use of a dental diagnostic terminology: A survey of dental care providers in a dental practice. J Am Dent Assoc 2015;146(6):390–7. [PubMed: 26025826]
27. Tokede O, White J, Stark P, Vaderhobli R, Walji M, Schoonheim-Klein M, et al. Assessing the use of a standardized dental diagnostic terminology in an electronic health record. Journal of Dental Education 2013;77(1):24–36. [PubMed: 23314462]
28. Obadan EM, Ramoni RB, Kalenderian E. Lessons learned from dental patient safety case reports. The Journal of the American Dental Association 2015;146(5):318–26.e2. [PubMed: 25925524]
29. National Coordinating Council for Medication Error Reporting and Prevention 2015 [Available from: http://www.nccmerp.org/medErrorCatIndex.html.

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Definition	*Harm* caused to the patient by dental care, regardless of whether it is associated with an error or is considered preventable	*Injury* resulting from a medical/ dental intervention was **not due to the underlying clinical condition of the patient**	*Physical harm* that is **moderate or severe** due to treatment within **a timeframe relevant to the clinical scenario**	*Same definition now* ***EXCLUDING***: ***Hazards or potential harm***; errors, negligence, blame, accusations, or malpractice are not AEs. Do not include ***omissions.*** ***Quality of Care*** issues in the absence of harm is not AE Harm may be temporary or permanent (E2, G2). ***Do not include E1, G1***. Only include moderate to severe harm. Does not matter if treatment occurred internal (at the institution) or externally (outside the institution). Repeated treatment attempts with ***poor prognosis (heroic dentistry) without harm*** is not an AE (e.g. repeated attempt to restore teeth). ***Loss of function in the absence of harm*** is not an AE Face validity needed. Outcomes not classified as AE because it is **a common occurrence as a result of disease or condition**, not unexpected, within a reasonable range of the standard care.	Same definition-specifications now at a category level (e.g. Pain managed by OTC is not an AE Pain requiring treatment at a emergency department is an AE).
Method	Individual Sites Review own charts for specified triggers	Expansion to additional triggers	Same triggers Different Data Site specific	Cases pulled from across all sites, reviewed by all clinicians	AEs considered by trigger type, consensus built across same dataset
Changes		New definition with consideration of patient characteristics	New definition with temporal constraints, degree of harm, and specified physical harm	Definition by exclusion. No longer necessary and sufficient conditions but requirement for the absence of potentially confounding factors	Classification specification considerations (e.g. pain requiring emergency care)

Figure 1:
Chart Review Process

Table 1:

Chart Review Process

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Definition	*Harm* caused to the patient by dental care, regardless of whether it is associated with an error or is considered preventable	*Injury* resulting from a medical/ dental intervention was **not due to the underlying clinical condition of the patient**	*Physical harm* that is **moderate or severe** due to treatment within **a timeframe relevant to the clinical scenario**	*Same definition now* ***EXCLUDING***: ***Hazards or potential harm***; errors, negligence, blame, accusations, or malpractice are not AEs. Do not include ***omissions.*** ***Quality of Care*** issues in the absence of harm is not AE Harm may be temporary or permanent (E2, G2). ***Do not include E1, G1***. Only include moderate to severe harm. Does not matter if treatment occurred internal (at the institution) or externally (outside the institution). Repeated treatment attempts with ***poor prognosis (heroic dentistry) without harm*** is not an AE (e.g. repeated attempt to restore teeth). ***Loss of function in the absence of harm*** is not an AE Face validity needed. Outcomes not classified as AE because it is **a common occurrence as a result of disease or condition**, not unexpected, within a reasonable range of the standard care.	Same definition- specifications now at a category level (e.g. Pain managed by OTC is not an AE. Pain requiring treatment at an emergency department is an AE).
Method	Individual Sites Review own charts for specified triggers	Expansion to additional triggers	Same triggers Different Data Site specific	Cases pulled from across all sites, reviewed by all clinicians	AEs considered by trigger type, consensus built across same dataset
Changes		New definition with consideration of patient characteristics	New definition with temporal constraints, degree of harm, and specified physical harm	Definition by exclusion. No longer necessary and sufficient conditions but requirement for the absence of potentially confounding factors	Classification specification considerations (e.g. pain requiring emergency care)

Table 2:

Inter-rater reliability

Pairwise Inter-Rater Reliability				
	Site 1	**Site2**	**Site 3**	**Site 4**
Site 1	1	0.04	−0.106	0.16
Site 2	**0.04**	1	0.059	0.21
Site 3	**−0.11**	**0.059**	1	0.21
Site 4	**0.16**	**0.208**	**0.206**	1