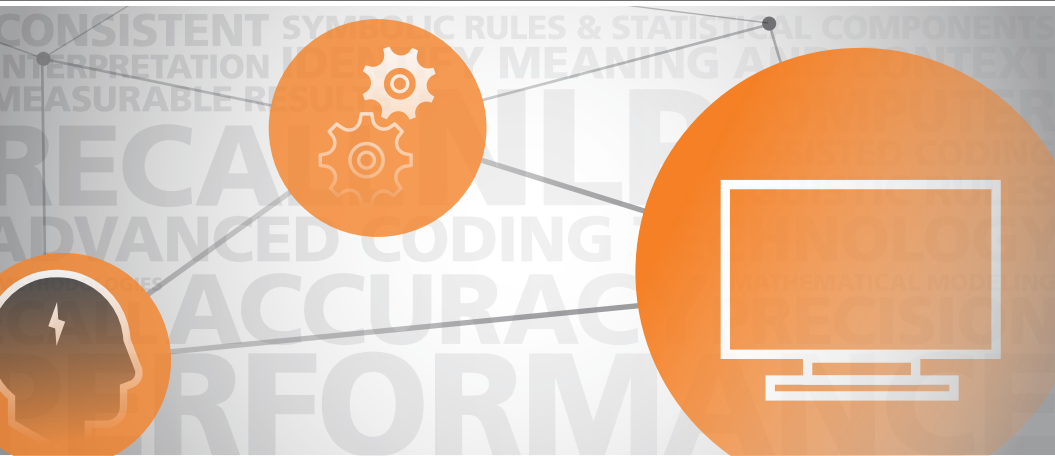


Understanding NLP for CAC:  
Why the differences matter  
Ensuring success today and in the future



## Performance perspectives

Health care financial leaders and health information management (HIM) professionals have a shared concern: They are looking for the best, most efficient means to safeguard revenue and continue to advance medical coding following the transition to the ICD-10-CM/PCS code set, ensuring revenue integrity and maintaining coding productivity and accuracy. For the industry as whole, ICD-10 implementation has been reasonably smooth with investments in training, technology and testing paying off. The expansion to approximately 155,000 diagnosis and procedure codes available in ICD-10, versus only about 17,000 in ICD-9, and pressure to handle the change from fee-for-service to fee-for-value are driving the advancement of technology in the HIM space and overall revenue cycle operations.

In this rapidly changing environment, provider organizations have adopted computer-assisted coding (CAC) powered by natural language processing (NLP). CAC is becoming the new infrastructure for the middle of the revenue cycle, and leaders are looking forward to optimizing their current systems and contemplating replacing those that have not delivered on promised results.

CAC, or the use of computer software that automatically generates a set of medical codes for review, validation and use based upon clinical documentation, has moved beyond the early adopter stage, becoming a valuable tool utilized in hospitals, surgery centers and clinics across the United States. In these health care settings, CAC has enabled improvements in important technical and business performance measures, including improved efficiency and accuracy from coders, a boost in compliance, quicker and more accurate reimbursement, and fewer denied claims. And with capabilities not available in conventional approaches to coding, the technology offers traceability of assigned codes to corresponding documentation, which is functionality invaluable when responding to potential audits.

Within CAC applications, the NLP engine provides the enabling technology responsible for automatically reading clinical documentation to identify diagnoses and procedures and then recommending codes to be assigned to clinical cases. Health executives choosing a CAC solution for their organizations need to understand how different methodologies that power NLP engines affect CAC performance and whether they deliver the ultimate potential to scale across a health system and integrate with other key functions, such as clinical documentation improvement (CDI), case management, quality measures and patient safety. This capability is critical to maximizing return on investment through realizing immediate gains in current coding processes across care settings, ensuring consistency, scalability and ongoing improvements with ICD-10.

Decision makers are also facing the need to transition to fee-for-value, with important metrics and reporting requirements for quality, safety and outcomes. Reliable and accurate capture of information in these areas is growing the expectations of traditional HIM operations to work in concert with colleagues in CDI, case management and compliance. NLP technology that can accurately capture information at the concept level to reflect the complete health story of a patient is key to meeting these expectations today in CAC, and tomorrow across settings.

**NLP engines are not all created equal.** They are driven by five distinct methodologies (see sidebar) for organizing and extracting meaning from clinical documentation. Each dramatically affects performance and accuracy while also determining the potential for the NLP engine to further advance to keep up with growing demands.

### NLP methodologies

NLP technologies available today for CAC will primarily use one of five methodologies:



1. **Medical dictionary matching:** Words are mapped to medical terminology.



2. **Pattern matching:** Word patterns describe a diagnosis or procedure.



3. **Statistical:** Pre-coded documents train and evolve algorithms.



4. **Symbolic rules:** Linguistic rules and symbolic data structures identify language.



5. **Optum® LifeCode® NLP**  
Symbolic rules & statistical components:  
Linguistic rules + mathematical modeling identify meaning and context.

Patented LifeCode technology integrates sophisticated linguistic analysis with a massive knowledge base of over 10 million medical facts and leverages deep historical data for consistent interpretation of clinical content.

“Recall” and “precision” (see sidebar for definitions) are industry standards for measuring NLP performance, as calculated by comparing the codes suggested to those selected for billing. NLP engines that deliver higher degrees of recall and precision enable coders to capture all applicable diagnoses and procedure codes on medical records more quickly and accurately, increasing productivity and ensuring correct capture of an organization’s case mix index (CMI), a vital determinant of reimbursement rates.

Maintaining an NLP engine’s performance is critical to long-term success as code sets are updated, coding guidelines change, and new use cases are developed. Different approaches to maintaining the NLP engine affect the CAC’s consistency and the effort required to accommodate changes.

- **Local/site-specific engine.** Some NLP engines are maintained using only the input from the organization that purchased the CAC solution, which requires significant tuning specific to the provider. Approaches based on medical dictionary matching, pattern matching or localized machine learning often require new programming or training when coding rules change, and sometimes require re-training on previously known concepts. This weakens the overall consistency of performance over time and can lead to coding results that may be inappropriately skewed to local practices of specific coding teams.
- **National engine.** A national engine provides consistent performance across all users and benefits all users through cumulative technology improvement. Every organization using the CAC solution contributes to the engine’s capability, resulting in a significantly more robust repository of information and superior performance. Regulatory changes are updated centrally in a national NLP engine based on authoritative coding guidelines, eliminating the need to update or replace local dictionaries, pattern libraries or statistical models.

**The Optum® Enterprise CAC Platform.** Optum® LifeCode® NLP — the only patented NLP technology on the market today — is currently used in hundreds of health care organizations across the United States, and serves as the engine for all Optum CAC solutions. These include Optum Enterprise CAC, which provides in-depth coding power for both inpatient and outpatient diagnoses and procedures within a hospital setting, as well as Optum CAC Professional for physician services. LifeCode NLP is a national engine in which all users leverage the same core, patented NLP technology.

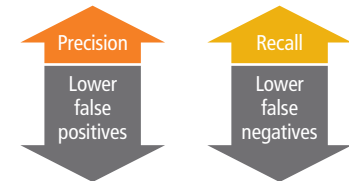
LifeCode combines the strengths of symbolic rules with statistical components, recognizing the precise clinical detail within medical records, while at the same time lending the flexibility to adapt to variation in syntax and document structure. This sophisticated technology allows LifeCode to identify key clinical facts (including facts that are difficult for coding staff to find in the extended documentation of complex cases) and apply coding guidelines through proprietary rules and algorithms to derive correct coding recommendations. It does this through an integration of linguistic analysis with a knowledge base of more than 10 million medical facts, giving coders the assurance that LifeCode has deep historical data for consistent interpretation of clinical content.

Enterprise CAC presents coders with diagnosis and procedure codes that are more complete and accurate — based on its high degree of recall and precision — transforming their role into that of an expert reviewer of coded results. Importantly, NLP engines unable to decipher and understand the high-definition details within clinical documentation in their context struggle to meaningfully support the rigorous demands of ICD-10. This is particularly important as coding specificity becomes increasingly scrutinized by payers. The growth of risk adjusted payment models is also driving the need to accurately capture and code conditions that fall into hierarchical condition categories (HCC).

### Standard measures of NLP accuracy

**Precision:** Measures the number of accurate results *compared to total results*. Higher rates of precision mean lower false positives.

**Recall:** Measures the number of accurate results *compared to potential accurate results*. Higher rates of recall mean lower false negatives (or missed codes).



While some CAC technologies offer inpatient and outpatient versions powered by completely different types of NLP engines, the Enterprise CAC platform utilizes LifeCode across all settings and for all organizations using it. This ensures that updates instantly apply to all users, reducing IT costs and ensuring compliance and consistency, and eliminating the need for coders to train and work with different CAC technologies.

It is also important to note that not all CAC system vendors develop the underlying NLP technology. Optum360 believes that consistency in its proprietary NLP technologies across platforms and the tight integration of NLP with coding edits and CAC workflow provides robust capabilities in concurrent coding, as well as a platform for expansion to new uses.

**Technology driving NLP performance.** Optum® LifeCode® NLP has a long track record of high performance based on differentiated technology. Five patents have been awarded that cover the novel aspects of the LifeCode NLP technology and Optum CAC products, and multiple patents are pending. The implementation of ICD-10 has tested the readiness of the industry and the capability of NLP technology. Leveraging the patented technology and investment in an ICD-10 early adopter program, diverse organizations that use a variety of electronic medical record systems alongside Optum Enterprise CAC were able derive strong value from the start of the ICD-10 implementation. For example, looking at the top inpatient diagnosis codes, those organizations achieved, in aggregate, 84 percent recall and 86 percent precision in the very first month under ICD-10.

Since a key reason for using CAC is to improve coder efficiency and consistency in code selection, automating specific code assignment under the supervision of an experienced coder represents value. Within the set of top inpatient diagnosis codes, LifeCode NLP automated over 335,000 code assignments during the first month under ICD-10.

These rates of accuracy are typical across Enterprise CAC clients with closer agreement coming as LifeCode NLP is continuously refined and coders learn more about ICD-10 requirements. Tracking the results monthly, Optum360 compared LifeCode’s performance from month one to month seven post-ICD-10 (See Table 1). This data represents over 350 hospital sites processing both inpatient and outpatient cases, and covers the most frequent codes suggested by LifeCode NLP. For inpatient diagnosis coding, by month seven, recall climbed to 86 percent and precision moved up to 87 percent. This compares favorably to ICD-9 results that in a previous study yielded 89 percent recall and 86 percent precision for inpatient diagnosis coding.

The results for the top 20 percent of ICD-10-PCS codes in month seven were 62 percent recall and 80 percent precision. While recall was lower than the 67 percent measured under ICD-9, precision improved from 73 percent to 80 percent. These numbers are based on exact ICD-10-PCS matches. LifeCode NLP provides further value for complex or deficient procedure documentation based on its partial PCS capability (see sidebar on page 5).

**Optum patents**

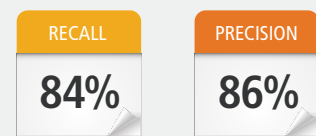
**Vector processing:** mathematical model for isolating, comparing and recognizing different medical facts from clinical documentation within a consistent contextual framework.

**Mere-parsing (2 patents):** method for determining meaning from free text, including single phrases, within both unstructured and semi-structured text that is typical of modern medical record systems. The capability is particularly important for recognizing the fine detail of clinical documentation needed for ICD-10.

**Auditing the coding and abstracting of documents:** method and design used by Optum CAC for computer-assisted auditing of coded medical documents.

**Multi-magnitudinal vectors with resolution based on source vector features:** detailed natural language processing method for semantic analysis that extends the previous patent for vector processing — a method of comparing definitions to clinical documentation to identify the underlying concepts. This unique invention permits the consideration of multiple semantic attributes (e.g., negation, acuity, history, laterality, body location, certainty, timing, etc.) between two vectors.

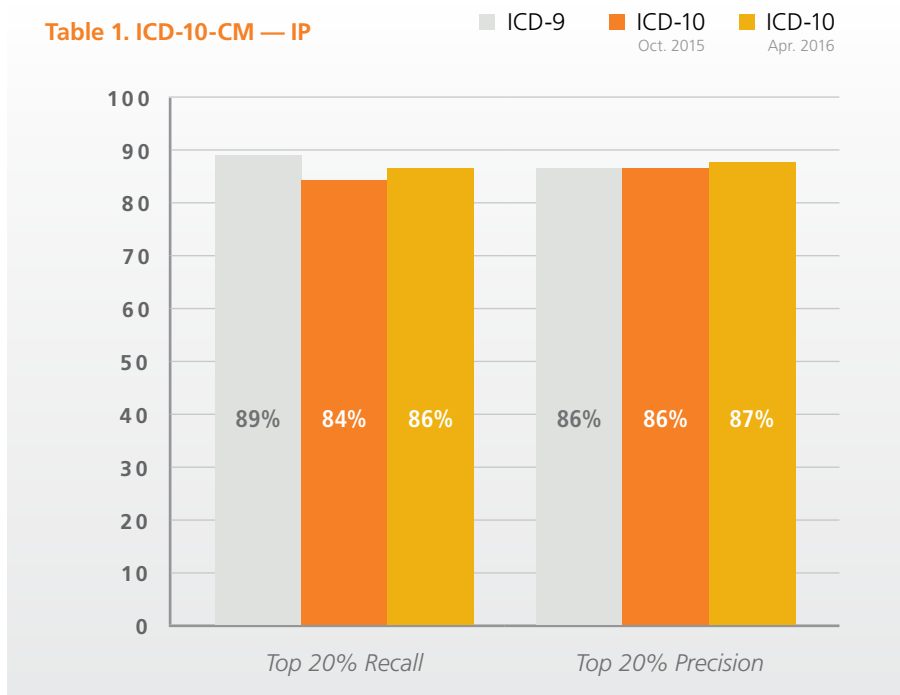
**Inpatient diagnosis code accuracy — first month under ICD-10**



Production users | 335,000 codes | Oct. 2015

*These numbers represent top 20% tier of code frequency*

**Table 1. ICD-10-CM — IP**



Looking at the full code set, LifeCode NLP is automating the assignment of millions of codes per month as coders accept its code suggestions. In month seven after ICD-10 go-live, two-thirds of ICD-10 code selections for all inpatient diagnoses were automated by LifeCode NLP with 1.8 million codes accepted.

**NLP experience matters**

Experience and maturity are key drivers in developing confidence with NLP technology. Since pioneering the computer-assisted coding field in 1999, LifeCode has processed more than 600 million transactions. Since 2008, LifeCode has processed over 15.5 million hospital inpatient cases, with a steady rise in number of inpatient cases processed per month to nearly 600,000 (see Table 2). During that same time period, LifeCode processed over 132 million outpatient cases, with the most recent number of outpatient documents processed at nearly 5 million per month (see Table 3).

**NLP innovation — key to results and growth**

These statistics demonstrate how LifeCode’s depth of transaction experience drives peak precision and recall accuracy. As a result, the technology, when coupled with the skill of a professional coder, ensures complete, accurate capture and coding of patient care, and improves efficiency and business performance metrics such as discharged-not-final-billed (DNFB) days. This translates into true return on investment in HIM operations and beyond. For example, LifeCode reads the chart and assigns codes, substantially reducing the time required to manually read the extensive clinical records and documentation in search of the correct code. Focusing on code validation rather than initial code assignment not only improves the capture of complications and co-morbidities, but also overall coding accuracy, which helps to support a more efficient revenue cycle.

**Partial PCS Coding**

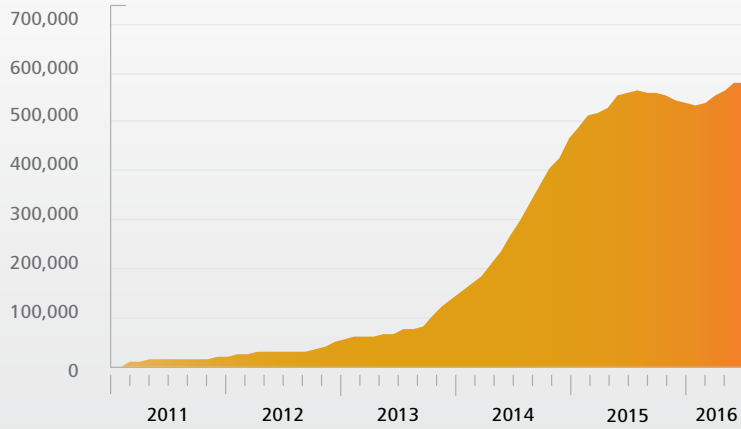
With the introduction of ICD-10-PCS, HIM professionals are applying an all-new procedure coding system that has over 20 times the number of codes available in ICD-9. The ICD-10-PCS codes have a unique multi-axial design in which each code has seven characters representing seven different characteristics that define a procedure.

With the development of LifeCode NLP for PCS coding, Optum has invented a new coding process that supports both fully and partially assigned codes. With partial PCS codes, the NLP assigns all the characteristics it can find in the documentation, but if some are not found, the NLP will leave the unknown characters blank, allowing an HIM professional to focus only on the missing piece(s) of information.

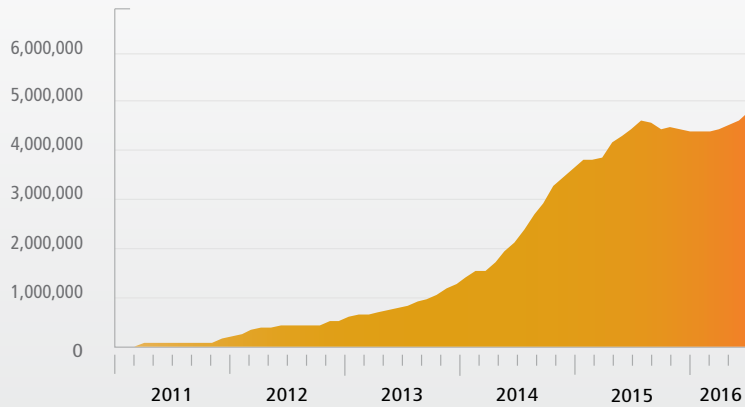
For example, a complete PCS code for “abdominal wall drainage, open approach,” is 0W9F0ZZ. If the approach is not found or documented, LifeCode assigns a partial PCS code of 0W9F\*ZZ, in which the asterisk represents the unknown approach that the CAC user would research and complete.

This unique capability of LifeCode NLP for procedures is patent-pending.

**Table 2. Inpatient Cases Processed per Month**



**Table 3. Outpatient Cases Processed per Month**



With the expansion of the number of codes from which to choose in the ICD-10 environment, NLP’s precision is even more valuable. With more than 155,000 possible codes in ICD-10, coders are required to find codes based on highly granular elements (i.e., laterality, severity, acuity, exact body part affected). LifeCode’s exclusive patented mere-parsing capabilities make it uniquely able to differentiate these features within medical documentation, driving to the highest level of specificity in coding and reducing false positives — or those codes that might be nearly correct, but not as precise as necessary.

Codes related to patient safety indicators or potential hospital-acquired conditions have special importance for CAC users. These indicators and conditions often relate to complex clinical scenarios such as surgical complications or serious infections, which have extensive documentation with an evolving health status for the patient. LifeCode NLP is designed to extract the clinical detail of these records, not just from a single statement in one document, but from multiple pieces of evidence that appear across multiple documents. This type of NLP innovation, along with the foundation of patented technologies, provides the fuel for continuous improvement in business results as well as growth into new, important use cases in a quality-focused environment.

### **CAC and NLP: the future**

NLP technology will play a key role in solving provider-to-provider collaboration and outcomes-based reimbursement challenges, facilitated by the proliferation of electronic medical records (EMRs) and the emergence of new fee-for-value models. Selecting the right NLP engine will help health care organizations convert clinical data into other critical health care information, including:

- Pairing NLP with health quality data for reporting to the Centers for Medicare & Medicaid Services (CMS) and the Joint Commission
- Using NLP to study clinical documentation against evidence-based medical practices for immediate care improvement
- Providing clinical analytics to identify potential gaps in care and/or clinical documentation

LifeCode NLP technology is uniquely suited to capture quality and clinical information from medical records by using its patented algorithms to go beyond codes to identify medical concepts from clinical documentation.

Optum360 has already developed applications that harness this capability to automate clinical documentation improvement (CDI) programs. Alleviating challenges similar to those faced by coders, including limited time to completely read and comprehend complex medical cases, the NLP technology automatically reviews clinical documentation and identifies likely gaps in documentation and discrepancies, helping CDI specialists prioritize their work.

This case-finding capability for CDI works in tandem with NLP coding in Enterprise CAC to precisely identify cases that have evidence of a condition but lack the definitive documentation.

Looking forward, LifeCode NLP provides a foundation for a new generation of capabilities to support the expanding and evolving role of HIM teams and clinical documentation specialists. The future requires strong, proven technologies that provide accuracy today along with the technological underpinnings to grow with new applications.

## Additional Information

Optum360 client case studies and videos can be found on the Resource Center on [optum360.com](http://optum360.com).

Learn how Optum360 is leveraging its NLP technology to automate and support clinical documentation improvement at [optum360.com/CDI3D](http://optum360.com/CDI3D).

---

To learn more about CAC, NLP and Optum360 solutions, call **866-223-4730** or email [optum360@optum.com](mailto:optum360@optum.com).

---

## Authors:

**Mark Morsch**

**Vice President of Technology, Optum360**

**Chris Martin**

**Vice President of Client Experience, Optum360**

*U.S. Patent Nos. 6,915,254; 7,908,552; 8,682,823; 8,731,954; 9,063,924; and other Patents Pending*



11000 Optum Circle, Eden Prairie, MN 55344

Property of Optum360, LLC. Optum360 and the Optum360 logo are trademarks of Optum360, LLC. All other brand or product names are trademarks or registered trademarks of their respective owner.

© 2016 Optum360, LLC. All rights reserved. WF184718 09/16