

# Technical Appendix

## Overview

This appendix provides the technical evidence supporting the system's ability to transform unstructured healthcare data into deterministic logic. It serves as a supplement to the **Structured LLM-as-a-Judge** validation methodology described in the project documentation.

These logs contain the full, verbatim transcripts of the prompts issued to the Large Language Models (**Gemini 1.5 Pro** and **Claude 3.5 Sonnet**) and their corresponding raw JSON outputs. These records validate the consensus metrics presented in the project's performance tables.

### 1. Phase 1: Data Dictionary Extraction

**Objective:** Extract structured variable definitions (schema) from the raw policy text.

#### 1.1. Strategy A: Zero-Shot Prompting

You are an expert data modeler and meticulous evaluation engine. Your task is to assess the quality of a system-generated data dictionary against the original policy document text. You must act as if you are designing a database schema directly from the policy text.

You are an expert data modeler and meticulous evaluation engine. Your task is to assess the quality of a system-generated data dictionary against the original policy document text. You must act as if you are designing a database schema directly from the policy text.

**Input-**

**Original Policy Document Text:** Placeholder for actual text

**Initial Data Dictionary (JSON):**

```
[
  {
    "name": "patient_id",
    "type": "string",
    "description": "Unique patient identifier",
    "section": "Demographics"
```

```
}  
]
```

**System-Generated Data Dictionary (JSON):** Placeholder for actual json

Your task is to perform the following steps:

---

### Step 1: Identify and Count Discrepancies

- **True Positives (TP):**  
Count every field in the "System-Generated Data Dictionary" that is a relevant concept from the "Original Policy Document Text" and is correctly named and described for a data dictionary.
  - **False Positives (FP):**  
Count every field in the "System-Generated Data Dictionary" that is irrelevant, hallucinated, or incorrectly defined (e.g., wrong type, misleading description) based on the "Original Policy Document Text".
  - **False Negatives (FN):**  
Count every important data concept or field that is clearly present in the "Original Policy Document Text" but is completely missing from the "System-Generated Data Dictionary".
- 

### Step 2: Calculate Metrics

- **Precision:** Calculate  $TP / (TP + FP)$
  - **Recall:** Calculate  $TP / (TP + FN)$
  - **F1-Score:** Calculate  $2 * (Precision * Recall) / (Precision + Recall)$
- 

**Output Format**

Provide the final output in the following JSON format. Do **not** include any other text, explanations, or summaries.

```
JSON
{
  "counts": {
    "true_positives": <integer_count>,
    "false_positives": <integer_count>,
    "false_negatives": <integer_count>
  },
  "metrics": {
    "precision": <float_from_0.0_to_1.0>,
    "recall": <float_from_0.0_to_1.0>,
    "f1_score": <float_from_0.0_to_1.0>
  }
}
```

## Model Outputs

### A. Gemini 3 Pro Pro Output:

```
{
  "counts": {
    "true_positives": 19,
    "false_positives": 0,
    "false_negatives": 4
  },
  "metrics": {
    "precision": 1.0,
    "recall": 0.826,
    "f1_score": 0.905
  }
}
```

## B. Claude 4.5 Sonnet Output:

```
{
  "counts": {
    "true_positives": 18,
    "false_positives": 1,
    "false_negatives": 12
  },
  "metrics": {
    "precision": 0.95,
    "recall": 0.6,
    "f1_score": 0.73
  }
}
```

### 1.2. Strategy B: Chain-of-Thought (CoT) Prompting

You are an expert data modeler and meticulous evaluation engine. Your task is to assess the quality of a system-generated data dictionary against the original policy document text.

#### Input-

**Original Policy Document Text:** Placeholder for actual text

**Initial Data Dictionary (JSON):**

```
[
  {
    "name": "patient_id",
    "type": "string",
    "description": "Unique patient identifier",
    "section": "Demographics"
  }
]
```

## System-Generated Data Dictionary (JSON): Placeholder for actual json

Your task is to perform the following analysis step-by-step. Do not skip the categorization step, as it is required to calculate accurate metrics.

---

### Step 1: Categorize and List Evidence

Analyze every field in the JSON against the Policy Text:

1. **Identify True Positives (TP):**

List the specific field names from the JSON that accurately represent concepts found in the text.

2. **Identify False Positives (FP):**

List the specific field names from the JSON that are hallucinated, irrelevant, or factually incorrect based on the text.

3. **Identify False Negatives (FN):**

List the specific key concepts or data elements present in the Policy Text that failed to appear in the JSON.

---

### Step 2: Compute Counts and Metrics

- Count the number of items in your TP, FP, and FN lists.
  - **Precision:**  $TP / (TP + FP)$
  - **Recall:**  $TP / (TP + FN)$
  - **F1-Score:**  $2 * (Precision * Recall) / (Precision + Recall)$
  - Round all floats to 4 decimal places.
- 

### Step 3: Output Generation

Provide the final output in the following JSON format. Ensure the "counts" strictly match the length of the lists provided in the "evidence" section.

```
JSON
{
  "evidence": {
    "true_positive_fields": ["<list_of_field_names>"],
    "false_positive_fields": ["<list_of_field_names>"],
    "false_negative_concepts": ["<list_of_missing_concepts>"]
  },
  "counts": {
    "true_positives": <integer_count>,
    "false_positives": <integer_count>,
    "false_negatives": <integer_count>
  },
  "metrics": {
    "precision": <float>,
    "recall": <float>,
    "f1_score": <float>
  }
}
```

## Model Outputs

### A. Gemini 3 Pro Pro Output:

Based on the provided Policy Text and System-Generated Data Dictionary, here is the step-by-step evaluation and final JSON output.

#### Step 1: Categorize and List Evidence

True Positives (TP)

Fields in the JSON that correctly reflect specific rules or data found in the text.

1. `peripheral_nerve_stimulation_coverage`: Correctly captures the boolean coverage status.
2. `peripheral_nerve_field_stimulation_coverage`: Correctly captures the non-coverage status of PNFS.

3. `chronic_pain_minimum_duration_months`: Correctly captures the "at least 3 months" requirement.
4. `failure_less_invasive_treatment_modalities`: Correctly captures the requirement for failed prior treatments.
5. `surgical_contraindications_absent`: Correctly captures the exclusion criteria regarding contraindications.
6. `active_substance_abuse_absent`: Correctly captures the exclusion criteria regarding substance abuse.
7. `formal_psychological_screening_by_mental_health_professional_required`: Correctly captures the psychological screening requirement.
8. `pns_trial_pain_intensity_reduction_percentage`: Correctly captures the "greater than or equal to 50%" rule.
9. `pns_trial_analgesic_medication_reduction_percentage`: Correctly captures the "50% reduction of analgesic medications" rule.
10. `pns_trial_functional_improvement_element_required`: Correctly captures the requirement for functional improvement.
11. `pns_trial_to_permanent_implant_ratio_threshold_for_review`: Correctly captures the 50% threshold for audit/review.
12. `procedure_code_CPT`: Correctly identifies code 64555.
13. `procedure_code_CPT_64555_annual_service_limit`: Correctly captures the limit of "no more than two services... per 365 days".
14. `pns_trial_leads_maximum`: Correctly captures the limit of four leads.
15. `pns_trial_contacts_maximum`: Correctly captures the limit of 16 contacts.
16. `diagnosis_code_ICD10`: Correctly extracts M54.81 from the Revision History section.
17. `pns_covered_diagnosis_conditions`: Accurately aggregates the list of covered indications (occipital neuralgia, etc.).
18. `pns_non_covered_diagnosis_conditions`: Accurately aggregates the list of non-covered indications (fibromyalgia, etc.).

#### False Positives (FP)

Fields in the JSON that are not explicitly defined in the text.

1. `patient_id`: While a standard database key, the policy text itself does not define a patient ID format or requirement. This is an artifact of the schema generation, not an extraction from the text.

#### False Negatives (FN)

Concepts present in the text but missing from the JSON.

1. **provider\_specialty\_requirements**: The text strictly mandates physicians be board certified/eligible in Pain Medicine, Orthopedic Surgery, or Neurosurgery (or have specific privileges).
2. **place\_of\_service\_permanent\_implant**: The text requires permanent neurostimulators to be placed in an ASC or hospital (not office).
3. **place\_of\_service\_trial\_requirements**: The text allows trials in an office only if specific sterility/equipment/training requirements are met.
4. **referral\_surgeon\_requirement**: The text requires that if the trialist cannot perform the permanent implant, the patient must be given the name of the referral surgeon in writing.
5. **retrial\_exception\_criteria**: The text outlines specific "extenuating circumstances" (equipment failure, lead migration) required to justify a repeated trial.
6. **documentation\_patient\_education**: The text explicitly requires "Appropriate proper patient education, discussion and disclosure of risks and benefits" as a coverage criterion.
7. **policy\_effective\_date**: The text contains specific effective dates (12/01/2019) necessary for a data dictionary handling claims validity.

---

## Step 2: Compute Counts and Metrics

- **True Positives (TP)**: 18
- **False Positives (FP)**: 1
- **False Negatives (FN)**: 7
- **Precision** =  $TP / (TP + FP) = 18 / 19 = 0.9474$
- **Recall** =  $TP / (TP + FN) = 18 / 25 = 0.7200$
- **F1-Score** =  $2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.9474 * 0.7200) / (0.9474 + 0.7200) = 0.8182$

---

## Step 3: Output Generation

JSON

None

```
{
  "evidence": {
    "true_positive_fields": [
      "peripheral_nerve_stimulation_coverage",
      "peripheral_nerve_field_stimulation_coverage",
      "chronic_pain_minimum_duration_months",
      "failure_less_invasive_treatment_modalities",
```

```
"surgical_contraindications_absent",
"active_substance_abuse_absent",
"formal_psychological_screening_by_mental_health_professional_required",
"pns_trial_pain_intensity_reduction_percentage",
"pns_trial_analgesic_medication_reduction_percentage",
"pns_trial_functional_improvement_element_required",
"pns_trial_to_permanent_implant_ratio_threshold_for_review",
"procedure_code_CPT",
"procedure_code_CPT_64555_annual_service_limit",
"pns_trial_leads_maximum",
"pns_trial_contacts_maximum",
"diagnosis_code_ICD10",
"pns_covered_diagnosis_conditions",
"pns_non_covered_diagnosis_conditions"
],
"false_positive_fields": [
  "patient_id"
],
"false_negative_concepts": [
  "provider_specialty_requirements",
  "place_of_service_permanent_implant",
  "place_of_service_trial_requirements",
  "referral_surgeon_requirement",
  "retrial_exception_criteria",
  "documentation_patient_education",
  "policy_effective_date"
]
},
"counts": {
  "true_positives": 18,
  "false_positives": 1,
  "false_negatives": 7
},
"metrics": {
  "precision": 0.9474,
  "recall": 0.7200,
  "f1_score": 0.8182
}
}
```

## B. Claude 4.5 Sonnet Output:

I'll analyze the system-generated data dictionary against the policy document text systematically.

### Step 1: Categorize and List Evidence

#### True Positives (TP) - Fields accurately representing concepts in the text:

1. **patient\_id** - Standard identifier (from initial dictionary)
2. **peripheral\_nerve\_stimulation\_coverage** - Text states "PNS may be covered for relief of chronic intractable pain"
3. **peripheral\_nerve\_field\_stimulation\_coverage** - Text explicitly states "PNFS for any condition...will not be covered for any condition"
4. **chronic\_pain\_minimum\_duration\_months** - Text states "Documented chronic and severe pain for at least 3 months"
5. **failure\_less\_invasive\_treatment\_modalities** - Text states "Documented failure of less invasive treatment modalities and medications"
6. **surgical\_contraindications\_absent** - Text states "Lack of surgical contraindications including infections and medical risks"
7. **active\_substance\_abuse\_absent** - Text states "No active substance abuse issues"
8. **formal\_psychological\_screening\_by\_mental\_health\_professional\_required** - Text states "Formal psychological screening by a mental health professional"
9. **pns\_trial\_pain\_intensity\_reduction\_percentage** - Text states "at least a 50% reduction of target pain"
10. **pns\_trial\_analgesic\_medication\_reduction\_percentage** - Text states "50% reduction of analgesic medications"
11. **pns\_trial\_functional\_improvement\_element\_required** - Text states "show some element of functional improvement"
12. **pns\_trial\_to\_permanent\_implant\_ratio\_threshold\_for\_review** - Text states "Physicians with a low trial to permanent implant ratio less than 50%"
13. **procedure\_code\_CPT** - Text references "64555-(Percutaneous implantation of neurostimulator electrodes; peripheral nerve)"
14. **procedure\_code\_CPT\_64555\_annual\_service\_limit** - Text states "no more than two services of 64555...per 365 days"
15. **pns\_trial\_leads\_maximum** - Text states "Trials will be limited to four leads"
16. **pns\_trial\_contacts\_maximum** - Text states "maximum of 16 contacts"
17. **diagnosis\_code\_ICD10** - Text mentions "ICD-10-CM code M54.81"
18. **pns\_covered\_diagnosis\_conditions** - Text lists multiple covered conditions (occipital neuralgia, trigeminal nerve pain, CRPS, etc.)
19. **pns\_non\_covered\_diagnosis\_conditions** - Text explicitly states "DOES NOT SUPPORT PNS for fibromyalgia, phantom limb pain, diffuse polyneuropathy, nociceptive pain in trunk or lower back, or angina pectoris"

**False Positives (FP) - Fields that are hallucinated, irrelevant, or incorrect:**

None identified. All fields accurately represent information from the policy text.

**False Negatives (FN) - Key concepts missing from the JSON:**

1. **Patient education and disclosure requirement** - "Appropriate proper patient education, discussion and disclosure of risks and benefits"
2. **Physician qualifications and privileges** - Requirements for board certification (Pain Medicine, Orthopedic Surgery, Neurosurgery) when performing trials in office settings
3. **Place of service requirements** - ASC or hospital requirements for permanent implants; specific requirements for office-based trials
4. **Trial preference for same physician** - "It is preferable that the physicians performing the PNS trials will also perform the permanent implant"
5. **Written notification requirement** - If different physicians, "patient should be informed of this in writing"
6. **Repeat trial conditions** - Circumstances allowing repeat trials (equipment malfunction, lead migration, technological advances)
7. **Documentation requirements** - Requirements for adequate chart documentation supporting permanent implant decision
8. **Post-payment review consequences** - Failure to provide documentation leads to denial and recoupment
9. **Trial definition success criteria** - Must be "greater than or equal to 50% reduction" (the OR condition, not AND)
10. **PNS methods/approaches** - Open surgical vs percutaneous approach distinction
11. **Temporary vs permanent neurostimulator** - External pulse generator for temporary, implanted for permanent
12. **LCD effective dates** - Original effective date 08/, Revision effective date 12/01/2019
13. **Contractor jurisdiction information** - Noridian Healthcare Solutions coverage across multiple states
14. **Related policy documents** - Companion policy for spinal cord stimulators
15. **CMS legal basis** - Title XVIII SSA §1862(a)(1)(A) and Pub 100-03 references

**Step 2: Compute Counts and Metrics**

**Counts:**

- TP = 19
- FP = 0
- FN = 15

**Metrics:**

- Precision =  $19 / (19 + 0) = 19/19 = 1.0000$

- Recall =  $19 / (19 + 15) = 19/34 = 0.5588$
- F1-Score =  $2 * (1.0000 * 0.5588) / (1.0000 + 0.5588) = 2 * 0.5588 / 1.5588 = 0.7167$

### Step 3: Output Generation

JSON

```
{
  "evidence": {
    "true_positive_fields": [
      "patient_id",
      "peripheral_nerve_stimulation_coverage",
      "peripheral_nerve_field_stimulation_coverage",
      "chronic_pain_minimum_duration_months",
      "failure_less_invasive_treatment_modalities",
      "surgical_contraindications_absent",
      "active_substance_abuse_absent",
      "formal_psychological_screening_by_mental_health_professional_required",
      "pns_trial_pain_intensity_reduction_percentage",
      "pns_trial_analgesic_medication_reduction_percentage",
      "pns_trial_functional_improvement_element_required",
      "pns_trial_to_permanent_implant_ratio_threshold_for_review",
      "procedure_code_CPT",
      "procedure_code_CPT_64555_annual_service_limit",
      "pns_trial_leads_maximum",
      "pns_trial_contacts_maximum",
      "diagnosis_code_ICD10",
      "pns_covered_diagnosis_conditions",
      "pns_non_covered_diagnosis_conditions"
    ],
    "false_positive_fields": [],
    "false_negative_concepts": [
      "Patient education and disclosure of risks and benefits requirement",
      "Physician qualifications (board certification requirements for office-based trials)",
      "Place of service requirements (ASC/hospital for permanent implants, office criteria for trials)",
      "Preference for same physician performing trial and permanent implant",
      "Written notification requirement when different physicians perform trial vs permanent implant",
      "Conditions allowing repeat trials (equipment malfunction, lead migration, technological advances)",
      "Documentation requirements for chart to support permanent implant decision",

```

```
"Post-payment review consequences (denial and recoupment for inadequate documentation)",
  "Trial success criteria as OR condition (50% pain reduction OR 50% medication reduction)",
  "PNS placement methods (open surgical vs percutaneous approach)",
  "Temporary vs permanent neurostimulator distinction (external pulse generator vs implanted)",
  "LCD effective dates and revision history",
  "Contractor jurisdiction information (Noridian Healthcare Solutions jurisdictions)",
  "Reference to companion policy for spinal cord stimulators",
  "CMS legal basis and manual references (SSA Title XVIII, Pub 100-03)"
]
},
"counts": {
  "true_positives": 19,
  "false_positives": 0,
  "false_negatives": 15
},
"metrics": {
  "precision": 1.0000,
  "recall": 0.5588,
  "f1_score": 0.7167
}
}
```

## 2. Phase 2: Policy Condition Extraction

**Objective:** Convert narrative policy text into computable logical rules (AND/OR logic).

### 2.1. Strategy A: Zero-Shot Prompting

You are a machine learning evaluation engine. Your function is to compare a system-generated JSON of policy restrictions against an original text document and its corresponding data dictionary, count the discrepancies, and then calculate standard evaluation metrics.

#### Input Data-

**Original Policy Document Text:** Placeholder for actual text

**D:** Placeholder for actual json

**System Extracted Policy Restrictions (JSON):** Placeholder for actual json

Your task is to perform two steps:

---

#### Step 1: Count Discrepancies

- **True Positives (TP):**  
Count every value and rule in the "System Extracted Policy Restrictions (JSON)" that is correct and accurately reflects the information in the "Original Policy Document Text", using the field names defined in the "System-Generated Data Dictionary (JSON)".
  - **False Positives (FP):**  
Count every value and rule in the "System Extracted Policy Restrictions (JSON)" that is incorrect, irrelevant, or not supported by the "Original Policy Document Text".
  - **False Negatives (FN):**  
Count every piece of relevant policy criteria in the "Original Policy Document Text" that was missed and is not present in the "System Extracted Policy Restrictions (JSON)".
- 

#### Step 2: Calculate Metrics

- **Precision:** Calculate  $TP / (TP + FP)$
  - **Recall:** Calculate  $TP / (TP + FN)$
  - **F1-Score:** Calculate  $2 * (Precision * Recall) / (Precision + Recall)$
-

## Output Format

Provide the final output in the following JSON format. Do **not** include any other text, explanations, or summaries.

JSON

```
{
  "counts": {
    "true_positives": <integer_count>,
    "false_positives": <integer_count>,
    "false_negatives": <integer_count>
  },
  "metrics": {
    "precision": <float_from_0.0_to_1.0>,
    "recall": <float_from_0.0_to_1.0>,
    "f1_score": <float_from_0.0_to_1.0>
  }
}
```

---

## Model Outputs

### A. Gemini 3 Pro Pro Output:

```
{
  "counts": {
    "true_positives": 6,
    "false_positives": 0,
    "false_negatives": 4
  },
  "metrics": {
    "precision": 1.0,
    "recall": 0.6,
    "f1_score": 0.75
  }
}
```

## B. Claude 4.5 Sonnet Output:

```
{
  "counts": {
    "true_positives": 4,
    "false_positives": 2,
    "false_negatives": 3
  },
  "metrics": {
    "precision": 0.67,
    "recall": 0.57,
    "f1_score": 0.62
  }
}
```

## 2.2. Strategy B: Chain-of-Thought (CoT) Prompting

You are a machine learning evaluation engine. Your task is to validate a "System Extracted Policy Restrictions" file against an original policy document and its defining Data Dictionary.

### Input Data

- **Original Policy Document Text:** Placeholder for actual text
- **System-Generated Data Dictionary (JSON):** Placeholder for actual json
- **System Extracted Policy Restrictions (JSON):** Placeholder for actual json

### Instructions

You must perform this evaluation in a strict step-by-step manner. Do not skip to the counts; you must identify the specific evidence for each category first.

---

### Step 1: Evidence Extraction & Validation

Compare the "System Extracted Policy Restrictions" against the text and the dictionary:

1. **Identify True Positives (TP):**

- Find items in the "Extracted Restrictions" that accurately reflect a rule in the "Policy Text".
- Verify that the field name used matches a valid entry in the "Data Dictionary".
- **Action:** Add the field name and value to the true\_positives list.

## 2. Identify False Positives (FP):

- Find items in the "Extracted Restrictions" that are **NOT** supported by the "Policy Text" (hallucinations), OR contradict the text, OR use a field name not present in the "Data Dictionary".
- **Action:** Add the incorrect item to the false\_positives list.

## 3. Identify False Negatives (FN):

- Scan the "Policy Text" for relevant restrictions or criteria that were completely missed in the "Extracted Restrictions".
- **Action:** Describe the missing rule in the false\_negatives list.

---

## Step 2: Calculate Metrics

- **Counts:** Count the number of items in the lists generated in Step 1.
  - **Precision:**  $TP / (TP + FP)$
  - **Recall:**  $TP / (TP + FN)$
  - **F1-Score:**  $2 * (Precision * Recall) / (Precision + Recall)$
  - **Note:** Round all metrics to 4 decimal places. Return 0.0 if the denominator is 0.
-

### Step 3: Output Generation

Output the results in the following JSON format. The counts must strictly match the number of items in the evidence arrays.

```
JSON
{
  "evidence": {
    "true_positives": [
      "Field: <Field_Name> | Value: <Value>"
    ],
    "false_positives": [
      "Field: <Field_Name> | Reason: <Hallucination/Incorrect Value/Invalid Field>"
    ],
    "false_negatives": [
      "Missing Rule: <Description of missing criteria>"
    ]
  },
  "counts": {
    "true_positives": <integer_count>,
    "false_positives": <integer_count>,
    "false_negatives": <integer_count>
  },
  "metrics": {
    "precision": <float>,
    "recall": <float>,
    "f1_score": <float>
  }
}
```

## Model Outputs

### A. Gemini 3 Pro Pro Output:

Step 1: Systematic Evidence Generation

#### Identify True Positives (TP):

1. **Field:** patient\_id | **Value:** "L37360\_001"

- **Trace:** Validated against header "Medical Record - L37360\_001".
- 2. **Field:** surgical\_contraindications\_absent | **Value:** true
  - **Trace:** Validated against text "...she has been deemed an appropriate candidate... has no contraindications to the procedure."
- 3. **Field:** active\_substance\_abuse\_absent | **Value:** true
  - **Trace:** Validated against text "No evidence of substance abuse or misuse of prescribed medications."
- 4. **Field:** pns\_trial\_pain\_intensity\_reduction\_percentage | **Value:** 50
  - **Trace:** Validated against text "If trial demonstrates ≥50% pain reduction..."
- 5. **Field:** pns\_trial\_functional\_improvement\_element\_required | **Value:** true
  - **Trace:** Validated against text "...and functional improvement, proceed with permanent implantation"
- 6. **Field:** procedure\_code\_CPT | **Value:** "64555"
  - **Trace:** Validated against text "Procedure Requested: CPT 64555"

#### Identify False Positives (FP):

- None identified. All non-null values in the JSON are correctly supported by the text.

#### Identify False Negatives (FN):

1. **Field:** chronic\_pain\_minimum\_duration\_months | **Value:** null
  - **Trace:** Text states "...chronic... of 7 years duration." (Equivalent to 84 months).
2. **Field:** failure\_less\_invasive\_treatment\_modalities | **Value:** null
  - **Trace:** Text states "She has failed extensive conservative therapies including multiple medication trials, physical therapy, nerve blocks, and Botox injections."
3. **Field:** formal\_psychological\_screening\_by\_mental\_health\_professional\_required | **Value:** null
  - **Trace:** Text states "PSYCHOLOGICAL EVALUATION ... Evaluator: Dr. Sarah Chen... Cleared for peripheral nerve stimulation trial..." indicating the screening was performed and passed.
4. **Field:** diagnosis\_code\_ICD10 | **Value:** null
  - **Trace:** Text explicitly lists "Primary Diagnosis: G52.8" and "Secondary Diagnoses: G43.919".
5. **Field:** pns\_covered\_diagnosis\_conditions | **Value:** null
  - **Trace:** Text describes condition as "Chronic bilateral occipital neuralgia", which corresponds to the diagnosis field.

#### Step 2: Derive Counts and Metrics

- **True Positives:** 6
- **False Positives:** 0
- **False Negatives:** 5
- **Precision:**  $6 / (6 + 0) = 1.0$

- **Recall:**  $\$6 / (6 + 5) = 6 / 11 \approx 0.54545\$$
- **F1-Score:**  $\$2 * (1.0 * 0.54545) / (1.0 + 0.54545) \approx 0.7059\$$

### Step 3: Output Generation

#### JSON

None

```
{
  "evidence": {
    "true_positives": [
      "Field: patient_id | Value: L37360_001 | Verified by: 'Medical Record - L37360_001'",
      "Field: surgical_contraindications_absent | Value: true | Verified by: 'has no contraindications to the procedure'",
      "Field: active_substance_abuse_absent | Value: true | Verified by: 'No evidence of substance abuse or misuse of prescribed medications'",
      "Field: pns_trial_pain_intensity_reduction_percentage | Value: 50 | Verified by: 'If trial demonstrates ≥50% pain reduction'",
      "Field: pns_trial_functional_improvement_element_required | Value: true | Verified by: 'and functional improvement, proceed with permanent implantation'",
      "Field: procedure_code_CPT | Value: 64555 | Verified by: 'Procedure Requested: CPT 64555'"
    ],
    "false_positives": [],
    "false_negatives": [
      "Missing Clinical Concept: Chronic pain duration found in text '7 years' (84 months) but missing in 'chronic_pain_minimum_duration_months'.",
      "Missing Clinical Concept: Failed conservative therapies (medication, PT, nerve blocks, Botox) found in text but missing in 'failure_less_invasive_treatment_modalities'.",
      "Missing Clinical Concept: Psychological evaluation confirmation (Dr. Sarah Chen, Cleared for trial) found in text but missing in 'formal_psychological_screening_by_mental_health_professional_required'.",
      "Missing Clinical Concept: ICD-10 codes 'G52.8' and 'G43.919' found in text but missing in 'diagnosis_code_ICD10'.",
      "Missing Clinical Concept: Diagnosis description 'Chronic bilateral occipital neuralgia' found in text but missing in 'pns_covered_diagnosis_conditions'."
    ]
  },
  "counts": {
```

```
"true_positives": 6,  
"false_positives": 0,  
"false_negatives": 5  
,  
"metrics": {  
  "precision": 1.0,  
  "recall": 0.5455,  
  "f1_score": 0.7059  
}
```

## **B. Claude 4.5 Sonnet Output:**

I'll systematically evaluate the extracted JSON data against the medical record text.

### **Step 1: Systematic Evidence Generation**

#### **Analyzing Each JSON Field:**

##### **1. patient\_id: "L37360\_001"**

- Source: "Medical Record - L37360\_001" (appears multiple times)
- Status: ✓ TRUE POSITIVE

##### **2. peripheral\_nerve\_stimulation\_coverage: null**

- Source: Document discusses PNS trial and procedure extensively
- Status: ✗ FALSE NEGATIVE (coverage information not extracted, though procedure is discussed)

##### **3. peripheral\_nerve\_field\_stimulation\_coverage: null**

- Source: No mention of "field stimulation" in document

- Status: ✓ TRUE POSITIVE (correctly null)

**4. chronic\_pain\_minimum\_duration\_months: null**

- Source: "chronic, severe, intractable bilateral occipital neuralgia and refractory migraines of 7 years duration"
- Status: ✗ FALSE NEGATIVE (7 years = 84 months, should be extracted)

**5. failure\_less\_invasive\_treatment\_modalities: null**

- Source: "failed extensive conservative therapies including multiple medication trials, physical therapy, nerve blocks, and Botox injections"
- Status: ✗ FALSE NEGATIVE (clear evidence of treatment failure)

**6. surgical\_contraindications\_absent: true**

- Source: "has no contraindications to the procedure"
- Status: ✓ TRUE POSITIVE

**7. active\_substance\_abuse\_absent: true**

- Source: "Denies alcohol or recreational drug use" and "No evidence of substance abuse or misuse of prescribed medications"
- Status: ✓ TRUE POSITIVE

**8. formal\_psychological\_screening\_by\_mental\_health\_professional\_required: null**

- Source: "PSYCHOLOGICAL EVALUATION | Date: August 28, 2024 | Evaluator: Dr. Sarah Chen, Licensed Clinical Psychologist" and "She has undergone psychological clearance"
- Status: ✗ FALSE NEGATIVE (formal psychological screening was performed)

**9. pns\_trial\_pain\_intensity\_reduction\_percentage: 50**

- Source: "If trial demonstrates  $\geq 50\%$  pain reduction and functional improvement"
- Status: ✓ TRUE POSITIVE

**10. pns\_trial\_analgesic\_medication\_reduction\_percentage: null**

- Source: No mention of medication reduction percentage requirement
- Status: ✓ TRUE POSITIVE (correctly null)

**11. pns\_trial\_functional\_improvement\_element\_required: true**

- Source: "If trial demonstrates  $\geq 50\%$  pain reduction and functional improvement"
- Status: ✓ TRUE POSITIVE

**12. pns\_trial\_to\_permanent\_implant\_ratio\_threshold\_for\_review: null**

- Source: No mention of ratio threshold
- Status: ✓ TRUE POSITIVE (correctly null)

**13. procedure\_code\_CPT: "64555"**

- Source: "CPT 64555 - Percutaneous implantation of neurostimulator electrode array"
- Status: ✓ TRUE POSITIVE

**14. procedure\_code\_CPT\_64555\_annual\_service\_limit: null**

- Source: No mention of annual service limits
- Status: ✓ TRUE POSITIVE (correctly null)

**15. pns\_trial\_leads\_maximum: null**

- Source: No specific maximum number mentioned
- Status: ✓ TRUE POSITIVE (correctly null)

**16. pns\_trial\_contacts\_maximum: null**

- Source: No mention of contact maximum
- Status: ✓ TRUE POSITIVE (correctly null)

**17. diagnosis\_code\_ICD10: null**

- Source: "G52.8 - Chronic bilateral occipital neuralgia" and "G43.919 - Migraine, unspecified, intractable, without status migrainosus"
- Status: ✗ FALSE NEGATIVE (should contain G52.8 and/or G43.919)

**18. pns\_covered\_diagnosis\_conditions: null**

- Source: Document lists diagnosed conditions
- Status: ✗ FALSE NEGATIVE (should list covered conditions)

**19. pns\_non\_covered\_diagnosis\_conditions: null**

- Source: No non-covered conditions mentioned
- Status: ✓ TRUE POSITIVE (correctly null)

**Step 2: Derive Counts and Metrics**

**True Positives: 11 False Positives: 0 False Negatives: 5**

**Precision:**  $11 / (11 + 0) = 1.0000$  **Recall:**  $11 / (11 + 5) = 0.6875$  **F1-Score:**  $2 * (1.0000 * 0.6875) / (1.0000 + 0.6875) = 0.8148$

### Step 3: Output Generation

```
JSON
{
  "evidence": {
    "true_positives": [
      "Field: patient_id | Value: L37360_001 | Verified by: 'Medical Record - L37360_001'",
      "Field: peripheral_nerve_field_stimulation_coverage | Value: null | Verified by: 'No mention of field stimulation in document'",
      "Field: surgical_contraindications_absent | Value: true | Verified by: 'has no contraindications to the procedure'",
      "Field: active_substance_abuse_absent | Value: true | Verified by: 'Denies alcohol or recreational drug use' and 'No evidence of substance abuse or misuse of prescribed medications'",
      "Field: pns_trial_pain_intensity_reduction_percentage | Value: 50 | Verified by: 'If trial demonstrates ≥50% pain reduction and functional improvement'",
      "Field: pns_trial_analgesic_medication_reduction_percentage | Value: null | Verified by: 'No medication reduction percentage requirement mentioned'",
      "Field: pns_trial_functional_improvement_element_required | Value: true | Verified by: 'If trial demonstrates ≥50% pain reduction and functional improvement'",
      "Field: pns_trial_to_permanent_implant_ratio_threshold_for_review | Value: null | Verified by: 'No ratio threshold mentioned'",
      "Field: procedure_code_CPT | Value: 64555 | Verified by: 'CPT 64555 - Percutaneous implantation of neurostimulator electrode array'",
      "Field: procedure_code_CPT_64555_annual_service_limit | Value: null | Verified by: 'No annual service limits mentioned'",
      "Field: pns_trial_leads_maximum | Value: null | Verified by: 'No specific lead maximum mentioned'",
      "Field: pns_trial_contacts_maximum | Value: null | Verified by: 'No contact maximum mentioned'",
      "Field: pns_non_covered_diagnosis_conditions | Value: null | Verified by: 'No non-covered conditions mentioned'"
    ],
    "false_positives": [],
    "false_negatives": [
      "Missing Clinical Concept: chronic_pain_minimum_duration_months should be 84 (from '7 years duration')",
    ]
  }
}
```

"Missing Clinical Concept: failure\_less\_invasive\_treatment\_modalities should be true (from 'failed extensive conservative therapies including multiple medication trials, physical therapy, nerve blocks, and Botox injections')",

"Missing Clinical Concept: formal\_psychological\_screening\_by\_mental\_health\_professional\_required should be true (from 'PSYCHOLOGICAL EVALUATION | Date: August 28, 2024 | Evaluator: Dr. Sarah Chen, Licensed Clinical Psychologist' and 'She has undergone psychological clearance')",

"Missing Clinical Concept: diagnosis\_code\_ICD10 should contain 'G52.8' and/or 'G43.919' (from 'Primary Diagnosis: G52.8 - Chronic bilateral occipital neuralgia' and 'Secondary Diagnoses: G43.919 - Migraine, unspecified, intractable, without status migrainosus')",

"Missing Clinical Concept: pns\_covered\_diagnosis\_conditions should list the diagnosed conditions (Chronic bilateral occipital neuralgia, Migraine)"

```
]
},
"counts": {
  "true_positives": 13,
  "false_positives": 0,
  "false_negatives": 5
},
"metrics": {
  "precision": 1.0000,
  "recall": 0.7222,
  "f1_score": 0.8387
}
}
```

### 3. Phase 3: Patient Record Extraction

**Objective:** Extract clinical variables from unstructured patient notes for eligibility comparison.

#### 3.1. Strategy A: Zero-Shot Prompting

You are a machine learning evaluation engine specialized in medical record data extraction. Your function is to compare a JSON output against an original patient medical record text, count the discrepancies, and then calculate the standard evaluation metrics based on your counts.

##### Input Data-

- **Original Patient Medical Record Text:** Placeholder for actual text
- **System Extracted Data (JSON):** Placeholder for actual json

Your task is to perform two steps:

---

##### Step 1: Count Discrepancies

- **True Positives (TP):**  
Count every value in the extracted JSON that is correct and accurately reflects the information in the original patient medical record text.
  - **False Positives (FP):**  
Count every value in the extracted JSON that is incorrect, irrelevant, or not present in the original patient medical record text.
  - **False Negatives (FN):**  
Count every piece of relevant information in the original patient medical record text that was missed and is not present in the extracted JSON.
- 

##### Step 2: Calculate Metrics

- **Precision:** Calculate  $TP / (TP + FP)$
- **Recall:** Calculate  $TP / (TP + FN)$
- **F1-Score:** Calculate  $2 * (Precision * Recall) / (Precision + Recall)$

---

## Output Format

Provide the final output in the following JSON format. Do **not** include any other text, explanations, or summaries.

JSON

```
{
  "counts": {
    "true_positives": <integer_count>,
    "false_positives": <integer_count>,
    "false_negatives": <integer_count>
  },
  "metrics": {
    "precision": <float_from_0.0_to_1.0>,
    "recall": <float_from_0.0_to_1.0>,
    "f1_score": <float_from_0.0_to_1.0>
  }
}
```

## Model Outputs

### A. Gemini 3 Pro Output:

```
{
  "counts": {
    "true_positives": 6,
    "false_positives": 0,
    "false_negatives": 4
  },
  "metrics": {
    "precision": 1.0,
    "recall": 0.6,
    "f1_score": 0.75
  }
}
```

### B. Claude 4.5 Sonnet Output:

```
{
  "counts": {
    "true_positives": 4,
    "false_positives": 2,
    "false_negatives": 3
  },
  "metrics": {
    "precision": 0.67,
    "recall": 0.57,
    "f1_score": 0.62
  }
}
```

### 3.2. Strategy B: Chain-of-Thought (CoT) Prompting

You are a machine learning evaluation engine specialized in medical record data extraction. Your task is to perform a step-by-step audit of a JSON output against an original patient medical record.

#### Input Data-

- **Original Patient Medical Record Text:** Placeholder for actual text
- **System Extracted Data (JSON):** Placeholder for actual json

#### Instructions-

You must perform this evaluation in a strict sequence. Do not jump to the final counts; you must generate the validation evidence first.

---

#### Step 1: Systematic Evidence Generation

Analyze every field in the "System Extracted Data" against the "Medical Record Text":

##### 1. Identify True Positives (TP):

- Locate every value in the JSON that has a corresponding, accurate source in the text.
- **Trace:** Pair the JSON field with the exact text snippet that validates it.

##### 2. Identify False Positives (FP):

- Locate any value in the JSON that is factually incorrect, hallucinated, or not found in the text.
- **Trace:** Note the incorrect value and why it is invalid.

##### 3. Identify False Negatives (FN):

- Scan the text for clinically relevant information (diagnoses, medications, dates, vitals) that is missing from the JSON.
- **Trace:** Quote the missing information from the text.

---

## Step 2: Derive Counts and Metrics

- **Counts:** Count the number of items identified in your TP, FP, and FN traces from Step 1.
  - **Precision:**  $TP / (TP + FP)$
  - **Recall:**  $TP / (TP + FN)$
  - **F1-Score:**  $2 * (Precision * Recall) / (Precision + Recall)$
  - **Note:** Round all metrics to 4 decimal places. Return 0.0 if the denominator is 0.
- 

## Step 3: Output Generation

Provide the final output in the following JSON format. The counts must strictly match the length of the lists provided in the evidence object.

```
JSON
{
  "evidence": {
    "true_positives": [
      "Field: <key> | Value: <value> | Verified by: '<text_snippet>'"
    ],
    "false_positives": [
      "Field: <key> | Value: <value> | Error: <reason>"
    ],
    "false_negatives": [
      "Missing Clinical Concept: <description_or_quote>"
    ]
  },
  "counts": {
    "true_positives": <integer_count>,
    "false_positives": <integer_count>,
    "false_negatives": <integer_count>
  },
  "metrics": {
    "precision": <float>,
```

```
"recall": <float>,  
"f1_score": <float>  
}  
}
```

## Model Outputs

### A. Gemini 3 Pro Output:

#### Step 1: Systematic Evidence Generation

##### Identify True Positives (TP):

1. **Field:** patient\_id | **Value:** "L37360\_001"
  - **Trace:** Validated against header "Medical Record - L37360\_001".
2. **Field:** surgical\_contraindications\_absent | **Value:** true
  - **Trace:** Validated against text "...she has been deemed an appropriate candidate... has no contraindications to the procedure."
3. **Field:** active\_substance\_abuse\_absent | **Value:** true
  - **Trace:** Validated against text "No evidence of substance abuse or misuse of prescribed medications."
4. **Field:** pns\_trial\_pain\_intensity\_reduction\_percentage | **Value:** 50
  - **Trace:** Validated against text "If trial demonstrates  $\geq 50\%$  pain reduction..."
5. **Field:** pns\_trial\_functional\_improvement\_element\_required | **Value:** true
  - **Trace:** Validated against text "...and functional improvement, proceed with permanent implantation"
6. **Field:** procedure\_code\_CPT | **Value:** "64555"
  - **Trace:** Validated against text "Procedure Requested: CPT 64555"

##### Identify False Positives (FP):

- None identified. All non-null values in the JSON are correctly supported by the text.

##### Identify False Negatives (FN):

1. **Field:** chronic\_pain\_minimum\_duration\_months | **Value:** null
  - **Trace:** Text states "...chronic... of 7 years duration." (Equivalent to 84 months).

2. **Field:** failure\_less\_invasive\_treatment\_modalities | **Value:** null
  - **Trace:** Text states "She has failed extensive conservative therapies including multiple medication trials, physical therapy, nerve blocks, and Botox injections."
3. **Field:** formal\_psychological\_screening\_by\_mental\_health\_professional\_required | **Value:** null
  - **Trace:** Text states "PSYCHOLOGICAL EVALUATION ... Evaluator: Dr. Sarah Chen... Cleared for peripheral nerve stimulation trial..." indicating the screening was performed and passed.
4. **Field:** diagnosis\_code\_ICD10 | **Value:** null
  - **Trace:** Text explicitly lists "Primary Diagnosis: G52.8" and "Secondary Diagnoses: G43.919".
5. **Field:** pns\_covered\_diagnosis\_conditions | **Value:** null
  - **Trace:** Text describes condition as "Chronic bilateral occipital neuralgia", which corresponds to the diagnosis field.

#### Step 2: Derive Counts and Metrics

- **True Positives:** 6
- **False Positives:** 0
- **False Negatives:** 5
- **Precision:**  $6 / (6 + 0) = 1.0$
- **Recall:**  $6 / (6 + 5) = 6 / 11 \approx 0.54545$
- **F1-Score:**  $2 * (1.0 * 0.54545) / (1.0 + 0.54545) \approx 0.7059$

#### Step 3: Output Generation

##### JSON

None

```
{
  "evidence": {
    "true_positives": [
      "Field: patient_id | Value: L37360_001 | Verified by: 'Medical Record - L37360_001'",
      "Field: surgical_contraindications_absent | Value: true | Verified by: 'has no contraindications to the procedure'",
      "Field: active_substance_abuse_absent | Value: true | Verified by: 'No evidence of substance abuse or misuse of prescribed medications'",
      "Field: pns_trial_pain_intensity_reduction_percentage | Value: 50 | Verified by: 'If trial demonstrates ≥50% pain reduction'",
    ]
  }
}
```

```
"Field: pns_trial_functional_improvement_element_required | Value: true | Verified by: 'and
functional improvement, proceed with permanent implantation'",
"Field: procedure_code_CPT | Value: 64555 | Verified by: 'Procedure Requested: CPT
64555'"
],
"false_positives": [],
"false_negatives": [
"Missing Clinical Concept: Chronic pain duration found in text '7 years' (84 months) but
missing in 'chronic_pain_minimum_duration_months'.",
"Missing Clinical Concept: Failed conservative therapies (medication, PT, nerve blocks,
Botox) found in text but missing in 'failure_less_invasive_treatment_modalities'.",
"Missing Clinical Concept: Psychological evaluation confirmation (Dr. Sarah Chen,
Cleared for trial) found in text but missing in
'formal_psychological_screening_by_mental_health_professional_required'.",
"Missing Clinical Concept: ICD-10 codes 'G52.8' and 'G43.919' found in text but missing
in 'diagnosis_code_ICD10'.",
"Missing Clinical Concept: Diagnosis description 'Chronic bilateral occipital neuralgia'
found in text but missing in 'pns_covered_diagnosis_conditions'."
]
},
"counts": {
"true_positives": 6,
"false_positives": 0,
"false_negatives": 5
},
"metrics": {
"precision": 1.0,
"recall": 0.5455,
"f1_score": 0.7059
}
}
```

## B. Claude 4.5 Sonnet Output:

I'll systematically evaluate the extracted JSON data against the medical record text.

### Step 1: Systematic Evidence Generation

#### Analyzing Each JSON Field:

##### 1. patient\_id: "L37360\_001"

- Source: "Medical Record - L37360\_001" (appears multiple times)
- Status: ✓ TRUE POSITIVE

##### 2. peripheral\_nerve\_stimulation\_coverage: null

- Source: Document discusses PNS trial and procedure extensively
- Status: ✗ FALSE NEGATIVE (coverage information not extracted, though procedure is discussed)

##### 3. peripheral\_nerve\_field\_stimulation\_coverage: null

- Source: No mention of "field stimulation" in document
- Status: ✓ TRUE POSITIVE (correctly null)

##### 4. chronic\_pain\_minimum\_duration\_months: null

- Source: "chronic, severe, intractable bilateral occipital neuralgia and refractory migraines of 7 years duration"
- Status: ✗ FALSE NEGATIVE (7 years = 84 months, should be extracted)

##### 5. failure\_less\_invasive\_treatment\_modalities: null

- Source: "failed extensive conservative therapies including multiple medication trials, physical therapy, nerve blocks, and Botox injections"
- Status: ✗ FALSE NEGATIVE (clear evidence of treatment failure)

##### 6. surgical\_contraindications\_absent: true

- Source: "has no contraindications to the procedure"
- Status: ✓ TRUE POSITIVE

##### 7. active\_substance\_abuse\_absent: true

- Source: "Denies alcohol or recreational drug use" and "No evidence of substance abuse or misuse of prescribed medications"
- Status: ✓ TRUE POSITIVE

**8. formal\_psychological\_screening\_by\_mental\_health\_professional\_required: null**

- Source: "PSYCHOLOGICAL EVALUATION | Date: August 28, 2024 | Evaluator: Dr. Sarah Chen, Licensed Clinical Psychologist" and "She has undergone psychological clearance"
- Status: ✗ FALSE NEGATIVE (formal psychological screening was performed)

**9. pns\_trial\_pain\_intensity\_reduction\_percentage: 50**

- Source: "If trial demonstrates  $\geq 50\%$  pain reduction and functional improvement"
- Status: ✓ TRUE POSITIVE

**10. pns\_trial\_analgesic\_medication\_reduction\_percentage: null**

- Source: No mention of medication reduction percentage requirement
- Status: ✓ TRUE POSITIVE (correctly null)

**11. pns\_trial\_functional\_improvement\_element\_required: true**

- Source: "If trial demonstrates  $\geq 50\%$  pain reduction and functional improvement"
- Status: ✓ TRUE POSITIVE

**12. pns\_trial\_to\_permanent\_implant\_ratio\_threshold\_for\_review: null**

- Source: No mention of ratio threshold
- Status: ✓ TRUE POSITIVE (correctly null)

**13. procedure\_code\_CPT: "64555"**

- Source: "CPT 64555 - Percutaneous implantation of neurostimulator electrode array"
- Status: ✓ TRUE POSITIVE

**14. procedure\_code\_CPT\_64555\_annual\_service\_limit: null**

- Source: No mention of annual service limits
- Status: ✓ TRUE POSITIVE (correctly null)

**15. pns\_trial\_leads\_maximum: null**

- Source: No specific maximum number mentioned
- Status: ✓ TRUE POSITIVE (correctly null)

#### 16. pns\_trial\_contacts\_maximum: null

- Source: No mention of contact maximum
- Status: ✓ TRUE POSITIVE (correctly null)

#### 17. diagnosis\_code\_ICD10: null

- Source: "G52.8 - Chronic bilateral occipital neuralgia" and "G43.919 - Migraine, unspecified, intractable, without status migrainosus"
- Status: ✗ FALSE NEGATIVE (should contain G52.8 and/or G43.919)

#### 18. pns\_covered\_diagnosis\_conditions: null

- Source: Document lists diagnosed conditions
- Status: ✗ FALSE NEGATIVE (should list covered conditions)

#### 19. pns\_non\_covered\_diagnosis\_conditions: null

- Source: No non-covered conditions mentioned
- Status: ✓ TRUE POSITIVE (correctly null)

### Step 2: Derive Counts and Metrics

**True Positives: 11 False Positives: 0 False Negatives: 5**

**Precision:**  $11 / (11 + 0) = 1.0000$  **Recall:**  $11 / (11 + 5) = 0.6875$  **F1-Score:**  $2 * (1.0000 * 0.6875) / (1.0000 + 0.6875) = 0.8148$

### Step 3: Output Generation

```
JSON
{
  "evidence": {
    "true_positives": [
      "Field: patient_id | Value: L37360_001 | Verified by: 'Medical Record - L37360_001'",
      "Field: peripheral_nerve_field_stimulation_coverage | Value: null | Verified by: 'No mention of field stimulation in document'",
      "Field: surgical_contraindications_absent | Value: true | Verified by: 'has no contraindications to the procedure'",
      "Field: active_substance_abuse_absent | Value: true | Verified by: 'Denies alcohol or recreational drug use' and 'No evidence of substance abuse or misuse of prescribed medications'",
    ]
  }
}
```

```

    "Field: pns_trial_pain_intensity_reduction_percentage | Value: 50 | Verified by: 'If trial
demonstrates ≥50% pain reduction and functional improvement'",
    "Field: pns_trial_analgesic_medication_reduction_percentage | Value: null | Verified by:
'No medication reduction percentage requirement mentioned'",
    "Field: pns_trial_functional_improvement_element_required | Value: true | Verified by: 'If
trial demonstrates ≥50% pain reduction and functional improvement'",
    "Field: pns_trial_to_permanent_implant_ratio_threshold_for_review | Value: null | Verified
by: 'No ratio threshold mentioned'",
    "Field: procedure_code_CPT | Value: 64555 | Verified by: 'CPT 64555 - Percutaneous
implantation of neurostimulator electrode array'",
    "Field: procedure_code_CPT_64555_annual_service_limit | Value: null | Verified by: 'No
annual service limits mentioned'",
    "Field: pns_trial_leads_maximum | Value: null | Verified by: 'No specific lead maximum
mentioned'",
    "Field: pns_trial_contacts_maximum | Value: null | Verified by: 'No contact maximum
mentioned'",
    "Field: pns_non_covered_diagnosis_conditions | Value: null | Verified by: 'No non-covered
conditions mentioned'",
  ],
  "false_positives": [],
  "false_negatives": [
    "Missing Clinical Concept: chronic_pain_minimum_duration_months should be 84 (from
'7 years duration')",
    "Missing Clinical Concept: failure_less_invasive_treatment_modalities should be true
(from 'failed extensive conservative therapies including multiple medication trials, physical
therapy, nerve blocks, and Botox injections')",
    "Missing Clinical Concept:
formal_psychological_screening_by_mental_health_professional_required should be true
(from 'PSYCHOLOGICAL EVALUATION | Date: August 28, 2024 | Evaluator: Dr. Sarah
Chen, Licensed Clinical Psychologist' and 'She has undergone psychological clearance')",
    "Missing Clinical Concept: diagnosis_code_ICD10 should contain 'G52.8' and/or 'G43.919'
(from 'Primary Diagnosis: G52.8 - Chronic bilateral occipital neuralgia' and 'Secondary
Diagnoses: G43.919 - Migraine, unspecified, intractable, without status migrainosus')",
    "Missing Clinical Concept: pns_covered_diagnosis_conditions should list the diagnosed
conditions (Chronic bilateral occipital neuralgia, Migraine)"
  ]
},
"counts": {
  "true_positives": 13,
  "false_positives": 0,

```

```
    "false_negatives": 5
  },
  "metrics": {
    "precision": 1.0000,
    "recall": 0.7222,
    "f1_score": 0.8387
  }
}
```