1. Appendix

1.1. Compute Environment

Experiments are performed in a local on-prem university compute environment using 24 Intel Xeon 2.70GHz CPU cores, 8 Nvidia V100 GPUs, 4 Nvidia A100 GPUs, and 1.48 TB of RAM. All compute environments supported HIPAA-compliant data protocols.

Table 1: Criteria-level breakdown for best model (GPT-4 using the *ACIN* prompt)

| Criteria	True Value	Prec.	Rec.	F1	Support
ABDOMINAL	MET	0.7097	0.7333	0.7213	30
ABDOMINAL	NOT MET	0.8545	0.8393	0.8468	56
ADVANCED-CAD	MET	0.8542	0.9111	0.8817	45
ADVANCED-CAD	NOT MET	0.8947	0.8293	0.8608	41
ALCOHOL-ABUSE	MET	0.5000	0.6667	0.5714	3
ALCOHOL-ABUSE	NOT MET	0.9878	0.9759	0.9818	83
ASP-FOR-MI	MET	0.8590	0.9853	0.9178	68
ASP-FOR-MI	NOT MET	0.8750	0.3889	0.5385	18
CREATININE	MET	0.8333	0.8333	0.8333	24
CREATININE	NOT MET	0.9355	0.9355	0.9355	62
DIETSUPP-2MOS	MET	0.9286	0.8864	0.9070	44
DIETSUPP-2MOS	NOT MET	0.8864	0.9286	0.9070	42
DRUG-ABUSE	MET	0.7500	1.0000	0.8571	3
DRUG-ABUSE	NOT MET	1.0000	0.9880	0.9939	83
ENGLISH	MET	0.9863	0.9863	0.9863	73
ENGLISH	NOT MET	0.9231	0.9231	0.9231	13
HBA1C	MET	0.9394	0.8857	0.9118	35
HBA1C	NOT MET	0.9245	0.9608	0.9423	51
KETO-1YR	MET	0.0000	0.0000	0.0000	0
KETO-1YR	NOT MET	1.0000	1.0000	1.0000	86
MAJOR-DIABETES	MET	0.9730	0.8372	0.9000	43
MAJOR-DIABETES	NOT MET	0.8571	0.9767	0.9130	43
MAKES-DECISIONS	MET	0.9651	1.0000	0.9822	83
MAKES-DECISIONS	NOT MET	0.0000	0.0000	0.0000	3
MI-6MOS	MET	1.0000	0.7500	0.8571	8
MI-6MOS	NOT MET	0.9750	1.0000	0.9873	78
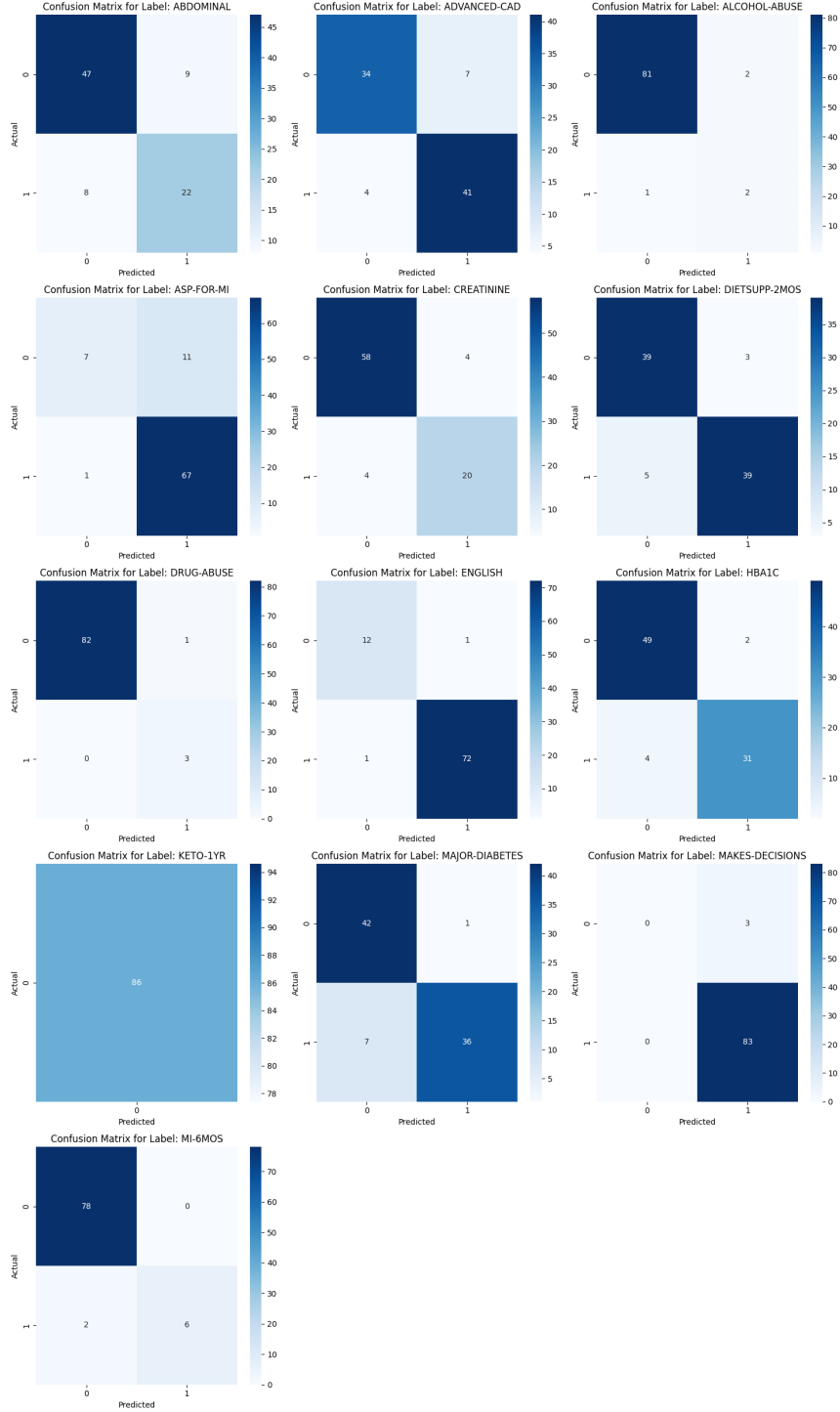
Figure 1: Confusion matrices for all criteria for best performing model (GPT-4 using the *ACIN* prompt).

Table 2: Eligibility criteria in the 2018 n2c2 cohort selection dataset, along with the percentage of patients who meet each criterion in the test set. Definitions taken verbatim from (**?**).

Criteria	Definition	Prevalence
ABDOMINAL	History of intra-abdominal surgery, small or large intestine resection, or small bowel obstruction	0.35
ADVANCED-CAD	Advanced cardiovascular disease (CAD). For the purposes of this annotation, we define "advanced" as having 2 or more of the following: ● Taking 2 or more medications to treat CAD ● History of myocardial infarction (MI) ● Currently experiencing angina ● Ischemia, past or present	0.52
ALCOHOL-ABUSE	Current alcohol use over weekly recommended limits	0.03
ASP-FOR-MI	Use of aspirin to prevent MI	0.79
CREATININE	Serum creatinine > upper limit of normal	0.28
DIETSUPP-2MOS	Taken a dietary supplement (excluding vitamin D) in the past 2 months	0.51
DRUG-ABUSE	Drug abuse, current or past	0.03
ENGLISH	Patient must speak English	0.85
HBA1C	Any hemoglobin A1c (HbA1c) value between 6.5% and 9.5%	0.41
KETO-1YR	Diagnosis of ketoacidosis in the past year	0.00
MAJOR-DIABETES	Major diabetes-related complication. For the purposes of this annotation, we define "major complication" (as opposed to "minor complication") as any of the following that are a result of (or strongly correlated with) uncontrolled diabetes: ● Amputation ● Kidney damage ● Skin conditions ● Retinopathy ● nephropathy ● neuropathy	0.50
MAKES-DECISIONS	Patient must make their own medical decisions	0.97
MI-6MOS	MI in the past 6 months	0.09

Table 3: "Improved" criteria definitions used in our prompts.

Criteria	Definition
ABDOMINAL	History of intra-abdominal surgery. This could include any form of intra-abdominal surgery, including but not limited to small/large intestine resection or small bowel obstruction
ADVANCED-CAD	Advanced cardiovascular disease (CAD). For the purposes of this annotation, we define "advanced" as having 2 or more of the following: (a) Taking 2 or more medications to treat CAD (b) History of myocardial infarction (MI) (c) Currently experiencing angina (d) Ischemia, past or present. The patient must have at least 2 of these categories (a,b,c,d) to meet this criterion, otherwise the patient does not meet this criterion. For ADVANCED-CAD, be strict in your evaluation of the patient – if they just have cardiovascular disease, then they do not meet this criterion.
ALCOHOL-ABUSE	Current alcohol use over weekly recommended limits
ASP-FOR-MI	Use of aspirin for preventing myocardial infarction (MI).
CREATININE	Serum creatinine level above the upper normal limit
DIETSUPP-2MOS	Consumption of a dietary supplement (excluding vitamin D) in the past 2 months. To assess this criterion, go through the list of medications_and_supplements taken from the note. If a substance could potentially be used as a dietary supplement (i.e. it is commonly used as a dietary supplement, even if it is not explicitly stated as being used as a dietary supplement), then the patient meets this criterion. Be lenient and broad in what is considered a dietary supplement. For example, a 'multivitamin' and 'calcium carbonate' should always be considered a dietary supplement if they are included in this list.
DRUG-ABUSE	Current or past history of drug abuse
ENGLISH	Patient speaks English. Assume that the patient speaks English, unless otherwise explicitly noted. If the patientś language is not mentioned in the note, then assume they speak English and thus meet this criteria.
HBA1C	Any hemoglobin A1c (HbA1c) value between 6.5% and 9.5%
KETO-1YR	Diagnosis of ketoacidosis within the past year
MAJOR-DIABETES	Major diabetes-related complication. Examples of "major complication" (as opposed to "minor complication") include, but are not limited to, any of the following that are a result of (or strongly correlated with) uncontrolled diabetes: • Amputation • Kidney damage • Skin conditions • Retinopathy • nephropathy • neuropathy. Additionally, if multiple conditions together imply a severe case of diabetes, then count that as a major complication.
MAKES-DECISIONS	Patient must make their own medical decisions. Assume that the patient makes their own medical decisions, unless otherwise explicitly noted. There is no information provided about the patientś ability to make their own medical decisions, then assume they do make their own decisions and therefore meet this criteria
MI-6MOS	Myocardial infarction (MI) within the past 6 months

```
# Task
Your job is to decide which of the following inclusion criteria the given patient meets.

# Patient

Below is a clinical note describing the patient's current health status:

```
{note}
```

Current Date

Assume that the current date is: {current_date}
Inclusion Criteria

The inclusion criteria being assessed are listed below, followed by their definitions:
{section_criteria}

Assessment

For each of the criteria above, use the patient's clinical note to determine whether the patient meets each criteria. Think step by step, and
justify your answer.

Format your response as a JSON list of dictionaries, where each dictionary contains the following elements:
* criterion: str - The name of the criterion being assessed
* medications_and_supplements: List[str] - The names of all current medications and supplements that the patient is taking
* rationale: str - Your reasoning as to why the patient does or does not meet that criterion
* is_met: bool - "true" if the patient meets that criterion, or it can be inferred that they meet that criterion with common sense. "false" if the
patient does not or it is impossible to assess this given the provided information.
* confidence: str - Either "low", "medium", or "high" to reflect your confidence in your response

An example of how your JSON response should be formatted is shown below, where the list of JSON dictionaries is stored in the "assessments"
key:
```json
{
  "assessments" : [
    {
      "criteria_1" : "something",
      "medications_and_supplements" : [ "medication_1", "medication_2"],
      "rationale" : "something something",
      "is_met" : true/false,
      "confidence" : "low/medium/high",
    },
    {
      "criteria_2" : "something",
      "medications_and_supplements" : [],
      "rationale" : "something something",
      "is_met" : true/false,
      "confidence" : "low/medium/high",
    },
    ...
  ]
}
```
The above example is only for illustration purposes only. It does not reflect the actual criteria or patient for this task.

Please analyze the given patient and inclusion criteria. Remember to include all inclusion criteria in your returned JSON dictionary. Please
provide your JSON response:
```

Figure 2: The prompt template we use for the **All Criteria** setting, i.e. the *ACIN* and *ACAN* strategies. The only variables injected into this prompt are **note**, **section_criteria**, and **current_date**. Respectively, these contain the text from the patient's EHR, the definitions of the inclusion criteria being evaluated, and the date of the last note in the patient's EHR

```
# Task
Your job is to decide whether the given patient meets the inclusion criterion.

# Patient

Below is a clinical note describing the patient's current health status:

```
{note}
```

Current Date

Assume that the current date is: {current_date}
Inclusion Criterion

The inclusion criterion being assessed is: "{criterion}: {definition}"

Assessment

Given the criterion above, use the patient's clinical note to determine whether the patient meets this criterion. Think step by step, and justify your answer.

Format your response as a JSON object with the following keys:
* criterion: str - The name of the criterion being assessed
* medications_and_supplements: List[str] - The names of all current medications and supplements that the patient is taking
* rationale: str - Your reasoning as to why the patient does or does not meet the criterion
* is_met: bool - "true" if the patient meets that criterion, or it can be inferred that they meet that criterion with common sense. "false" if the patient does not or it is impossible to assess this given the provided information.
* confidence: str - Either "low", "medium", or "high" to reflect your confidence in your response

An example of how your JSON response should be formatted is shown below:
```json
{
   "criterion" : "{criterion}",
   "medications_and_supplements" : [ "medication_1", "medication_2"],
   "rationale" : "something something",
   "is_met" : true/false,
   "confidence" : "low/medium/high",
}
```
The above example is only for illustration purposes only. It does not reflect the actual criterion or patient for this task.

Please provide your response:
```

Figure 3: The prompt template we use for the **Individual Criteria** setting, i.e. the *ICIN* and *ICAN* strategies. The only variables injected into this prompt are **note**, **section_criteria**, and **current_date**. Respectively, these contain the text from the patient's EHR, the definitions of the inclusion criteria being evaluated, and the date of the last note in the patient's EHR.