Large language models accurately identify people who inject drugs in Australian electronic health records

David Goodman-Meza, MD, PhD

Kirby Institute

St Vincent's Hospital Sydney

UNSW SYDNEY

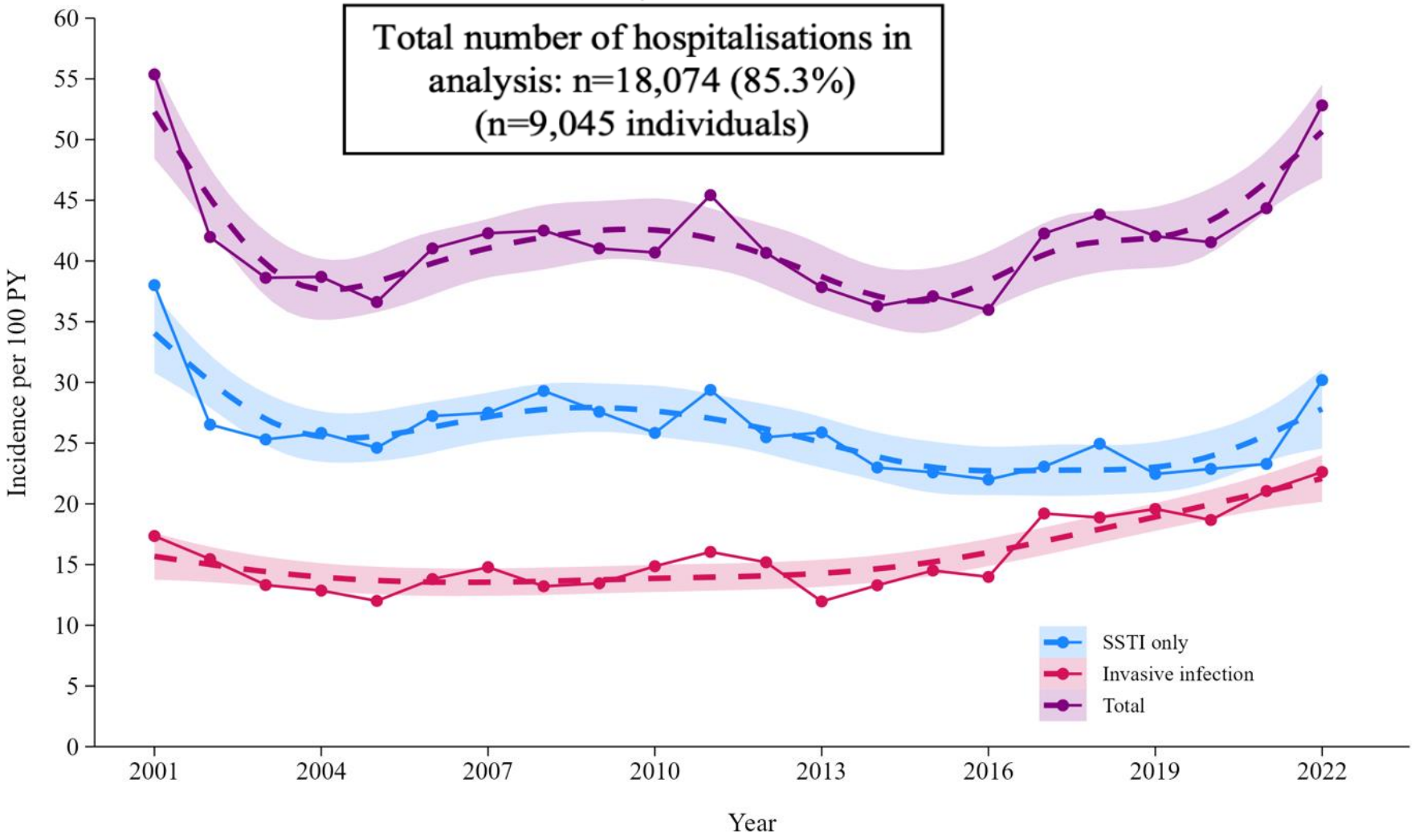
UNSW Kirby Institute

ST VINCENT'S
Better and fairer care. Always.

Acknowledgements

- Country
- Community
- Co-authors: Marianne Martinello, Jeff Masters, Denisse Russell, Brendan Jacka, Gail V Matthews, Greg J Dore
- No conflicts of interest to disclose

People Who Inject Drugs Are At High Risk For Serious Infections



Total

Skin/soft tissue

Invasive infection:
Endocarditis,
osteomyelitis
joint infections,
CNS

Masters J, et al. Open Forum Infect Dis. 2025

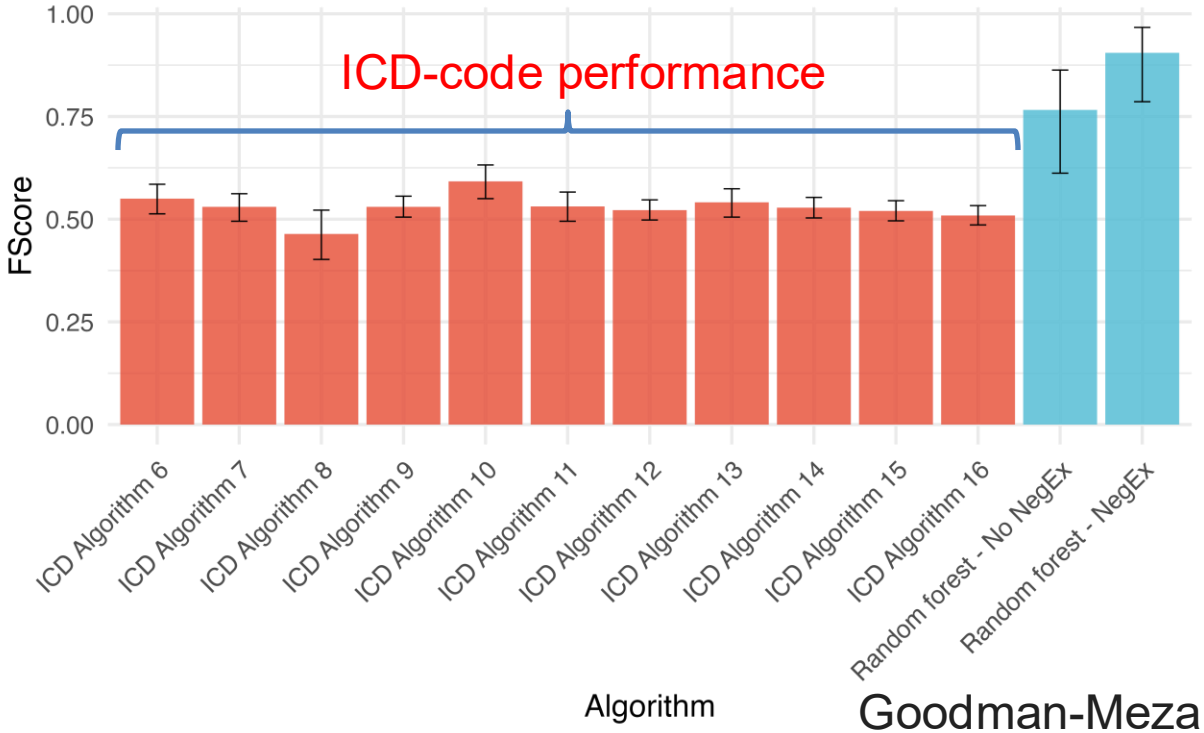
UNSW SYDNEY

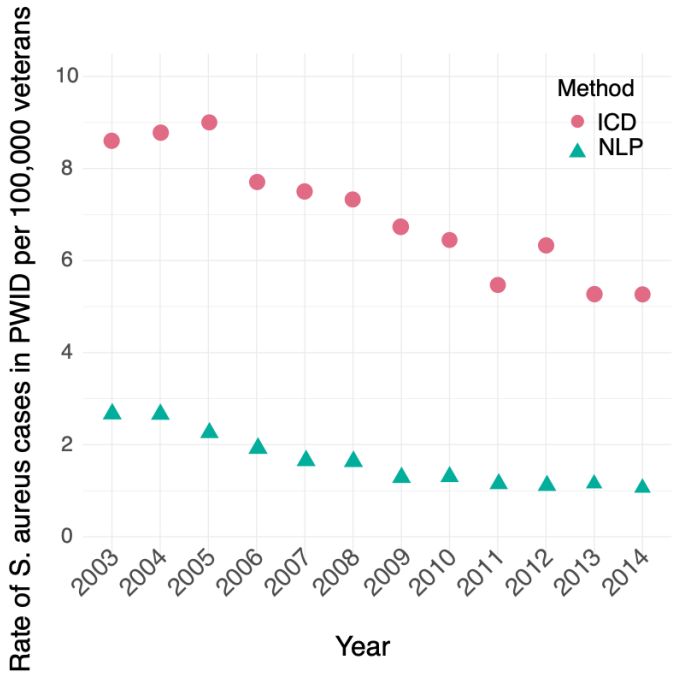
UNSW Kirby Institute

ST VINCENT'S Better and fairer care. Always.

Hospital Systems Lack Accurate Ways of Estimating the Care they Provide to People who Inject Drugs

- Existing methods rely on International Classification of Diseases codes (ICD)-10
  - No specific code identifies injection drug use
  - Proxy combinations: HIV, hepatitis C, substance use disorder, homelessness



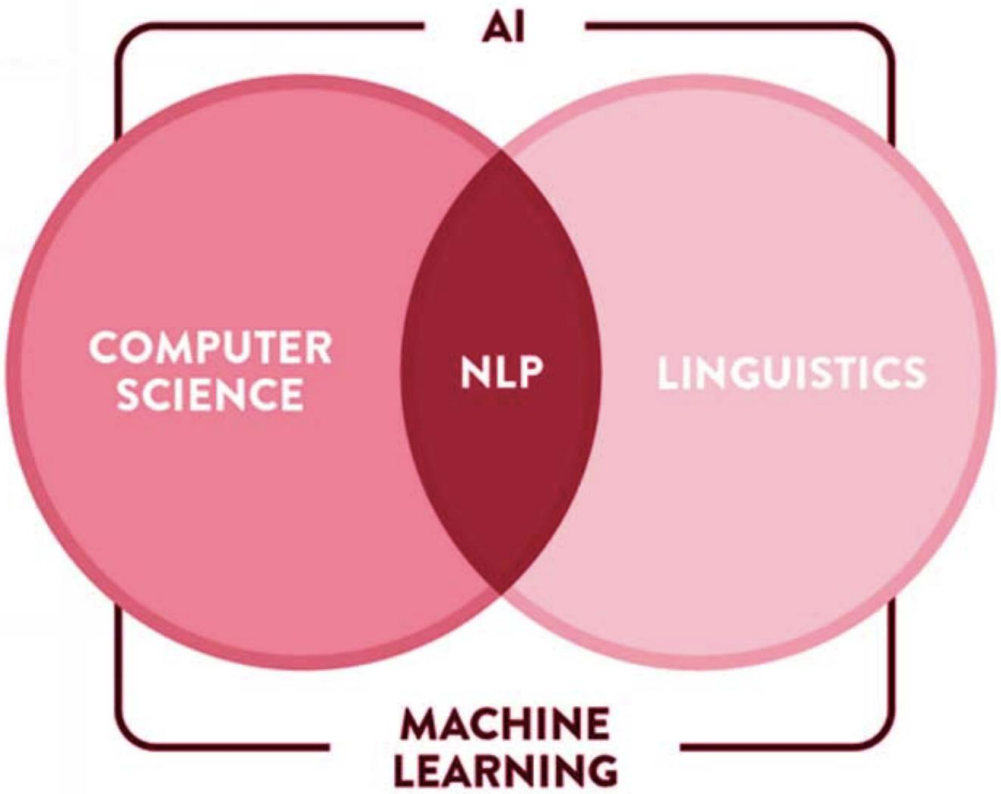


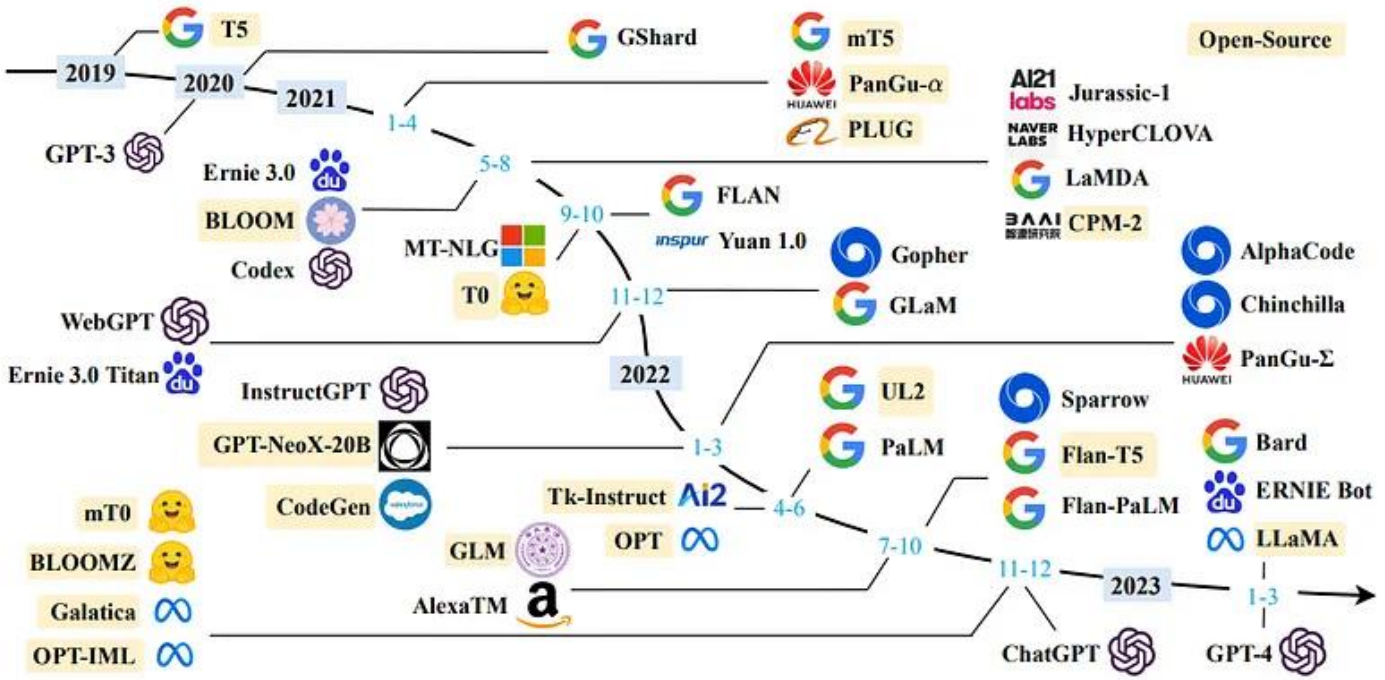
ICD-code estimate
7.04 per 100,000

NLP estimate
1.68 per 100,000

Goodman-Meza D, et al. Open Forum Infect Dis. 2024

New Artificial Intelligence Models May Improve Identification Of People Who Inject Drugs





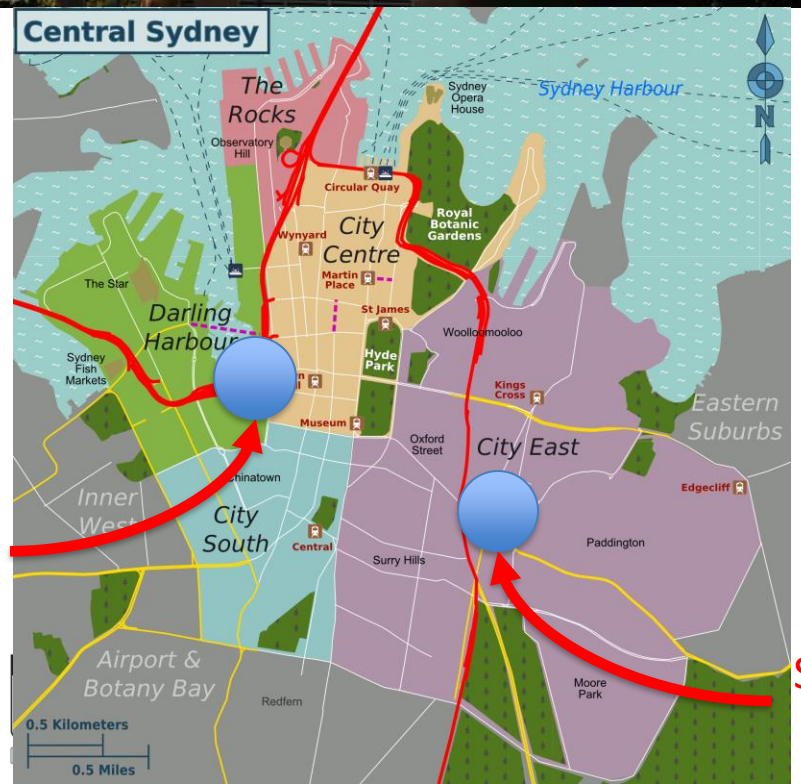
UNSW SYDNEY

UNSW Kirby Institute

ST VINCENT'S Better and fairer care. Always.

Methods

- Data:
  - St Vincent's Hospital Public Sydney
  - 2018–2022
  - Infectious Diseases service
  - Discharge summaries
  - De-identified



You Are Here

St Vincent's Hospital

Annotation (Gold Standard)

- **Prespecified annotator guide adapted to the Australian context**
- **Injection drug use**
  - Current/historical
- **Substances**
  - Heroin
  - Prescription opioids (misuse)
  - Fentanyl (misuse)
  - Methamphetamine
  - Cocaine
  - Cannabis
- **Opioid agonist treatment**
  - Methadone
  - Buprenorphine

- Single annotator

Variable	Values	Description
Injection drug use (dichotomous)	Yes	There is a mention that the case injected drugs or has a history of injection drug use. Terms may include: injection illicit drug use, injection ____ [specify drug], [specify drug] injector, IVDU, IVDA, PWID, IDU, intravenous drug user, intravenous drug abuser, injection drug user, skin popping, muscle popping. Key areas to look are in the History of Present Illness (HPI), social history, physical exam, and assessment and plan.
	No	There is no mention that the case injected drugs.
Chronicity (categorical)	Current:	Note reports that the case has injected an illicit substance within the last 30 days
	Historical:	Note reports that the case has injected an illicit substance more than 30 days ago or has a history of using a injecting a drug without a current mention
Type (categorical, multiple selection)	Heroin	Specified use of "heroin". Other terms include "black tar", "powder heroin", "china white", "dope", "smack", "horse".
	Prescription opioid	Specified illicit use of a prescription opioid. Examples include oxycodone, oxymorphone, hydrocodone, hydromorphone, Dilaudid, Opana, Norco, Percocet, Vicodin, oxys, methadone, buprenorphine. Fentanyl use should be classified with its own label.
	Fentanyl	Specified injection, smoking or snorting of "fentanyl".
	Cocaine or crack	Specified injection, smoking, or snorting of "cocaine", "crack" or "crack cocaine".
	Methamphetamine /amphetamine	Specified injection, smoking, or snorting of methamphetamine or amphetamines. Example terms include "meth", "methamphetamine", "amphetamine", "speed", "ice", "crystal". Notes may specify that the case used methamphetamine but not injected it.
	Benzodiazepines	Specified illicit use of benzos, benzodiazepines, diazepam, valium, alprazolam, zanax, clonazepam, konipin
	Cannabis/ marijuana	Specified use of cannabis, marijuana, THC, pot.

Comparators

ICD-10 codes

- Injection drug use combination
  - HIV (B21 to B24, Z21)
  - Hepatitis C (B17.1, B18.2, Z22.52)
  - Substance use related (Z72.2, Z86.41, T40.1 to T40.5, T43.6)
  - Substance-use disorders (F11, F14, F15, F19)
  - Homelessness (Z59)
- Specific substances:
  - Heroin (T40.1, F11)
  - Prescription opioids (T40.2 to T40.4, F11)
  - Methamphetamine (T43.6, F15).

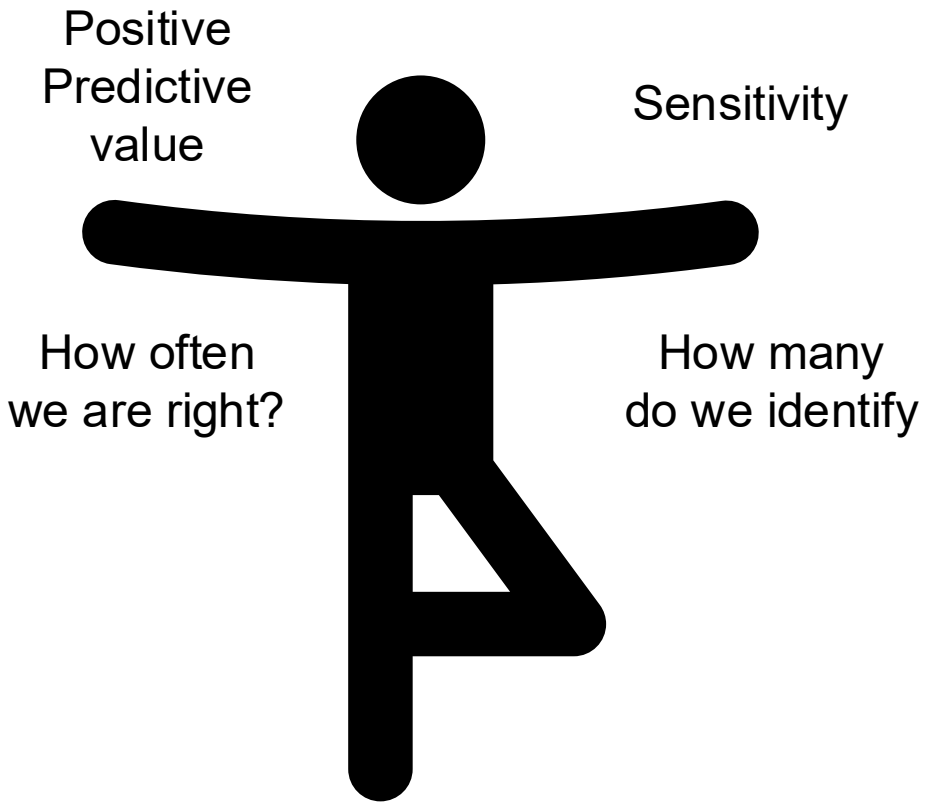
Large language models
All local via Ollama

- Llama 3.3 70 billion
- llama3-med42-70b
- llama3-med42-8b
- OpenBioLLM-Llama3-70B
- mistral
- phi4
- gemma3
- hippomistral

UNSW SYDNEY
UNSW Kirby Institute
ST VINCENT'S Better and fairer care. Always.

Statistical Analysis

- Diagnostic metrics
  - **F1 score**
    - **Balance of PPV and sensitivity)**
  - Accuracy
  - Sensitivity/specificity
  - Positive/negative predictive value



UNSW SYDNEY
UNSW Kirby Institute
ST VINCENT'S Better and fairer care. Always.

Results

859 Charts reviewed

149 (17.3%) Identified as
People who Inject Drugs

710 (82.6%) Identified as
People who do not Inject Drugs

UNSW SYDNEY

UNSW Kirby Institute

ST VINCENT'S
Better and fairer care. Always.

Results: Demographics

	People who inject drugs n = 149	People who do NOT inject drugs n = 710	P value
Age	45 (36, 50)	54 (39, 68)	<0.001

UNSW SYDNEY

UNSW Kirby Institute

ST VINCENT'S Better and fairer care. Always.

Results: Demographics

	People who inject drugs n = 149	People who do NOT inject drugs n = 710	P value
Age	45 (36, 50)	54 (39, 68)	<0.001
Female	32%	32%	NS

UNSW SYDNEY
UNSW Kirby Institute
ST VINCENT'S Better and fairer care. Always.

Results: Demographics

	People who inject drugs n = 149	People who do NOT inject drugs n = 710	P value
Age	45 (36, 50)	54 (39, 68)	<0.001
Female	32%	32%	NS
Identified as Aboriginal or Torres Strait Islander	23%	6%	<0.001

UNSW SYDNEY

UNSW Kirby Institute

ST VINCENT'S Better and fairer care. Always.

Results: Demographics

	People who inject drugs n = 149	People who do NOT inject drugs n = 710	P value
Age	45 (36, 50)	54 (39, 68)	<0.001
Female	32%	32%	NS
Identified as Aboriginal or Torres Strait Islander	23%	6%	<0.001
Length of stay, days	6 (3, 13)	4 (2, 8)	<0.001

UNSW SYDNEY

UNSW Kirby Institute

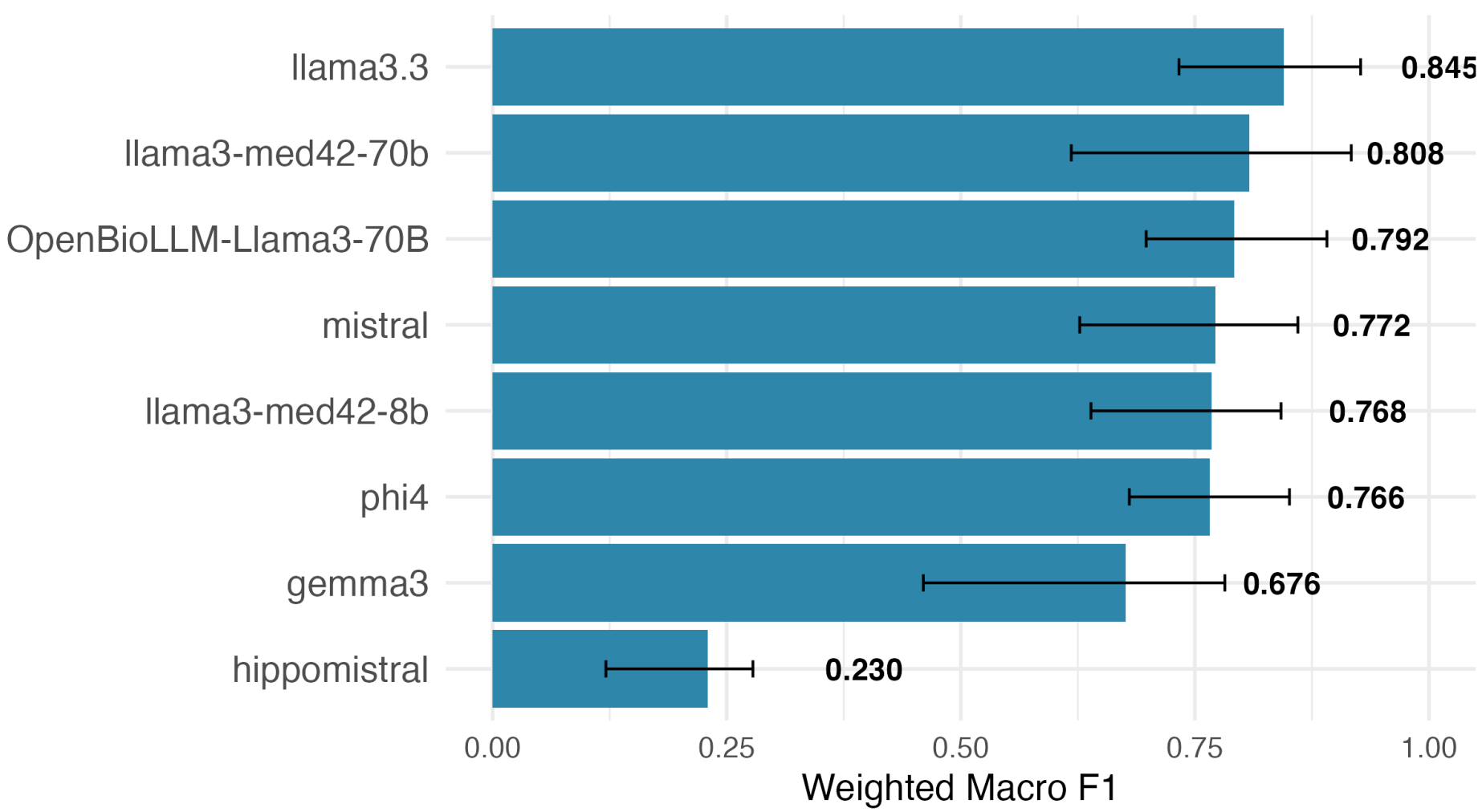
ST VINCENT'S Better and fairer care. Always.

Results: Demographics

	People who inject drugs n = 149	People who do NOT inject drugs n = 710	P value
Age	45 (36, 50)	54 (39, 68)	<0.001
Female	32%	32%	NS
Identified as Aboriginal or Torres Strait Islander	23%	6%	<0.001
Length of stay, days	6 (3, 13)	4 (2, 8)	<0.001
Patient directed discharge	22%	5%	<0.001

UNSW SYDNEY
UNSW Kirby Institute
ST VINCENT'S Better and fairer care. Always.

Llama3.3 Outperformed All Other LLMs

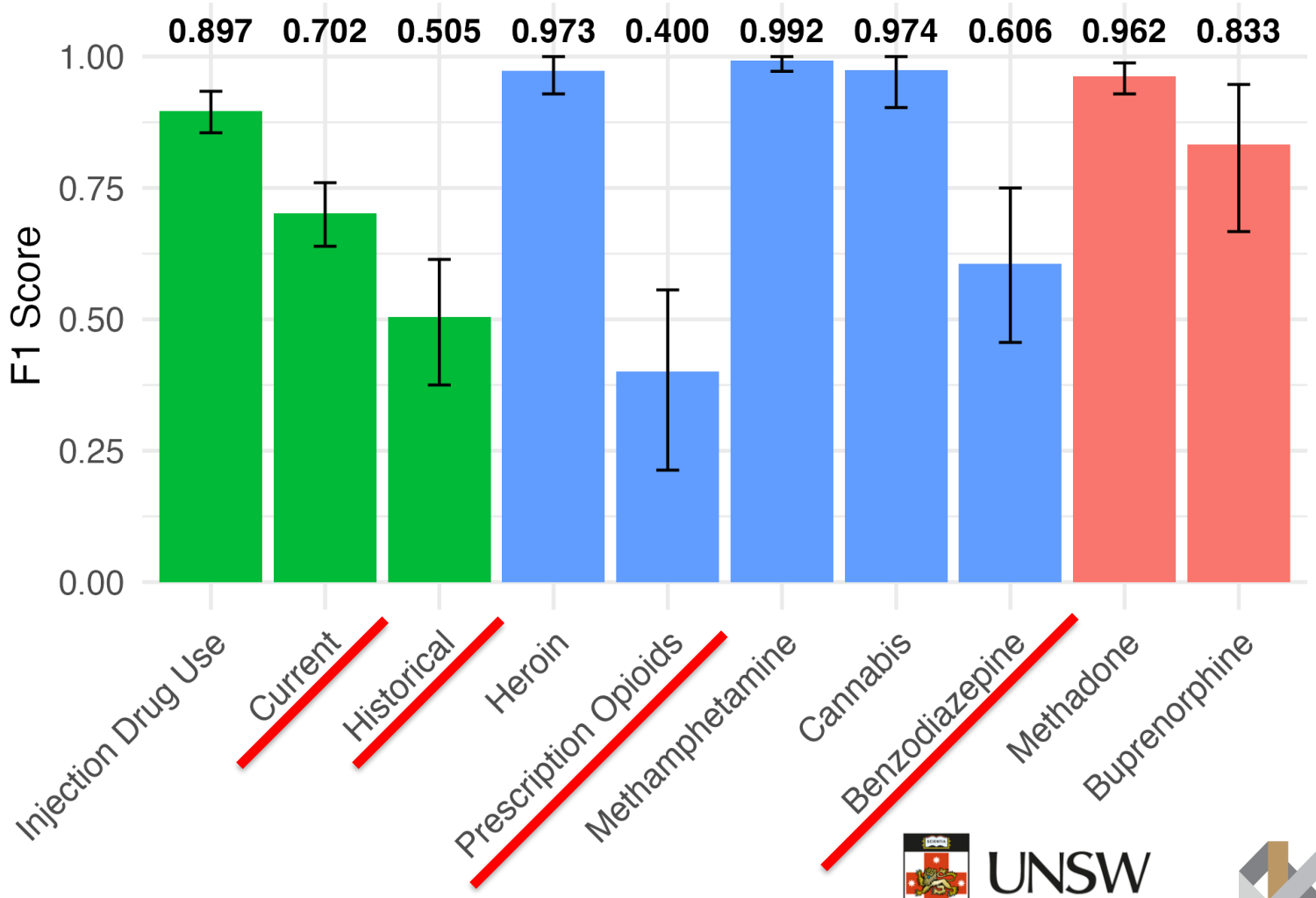


Sensitivity - 0.820
Specificity - 0.999

Error bars denote
95% confidence intervals

UNSW SYDNEY
UNSW Kirby Institute
ST VINCENT'S Better and fairer care. Always.

Poor LLM Performance For Acuity, Prescription Opioids And Benzodiazepines



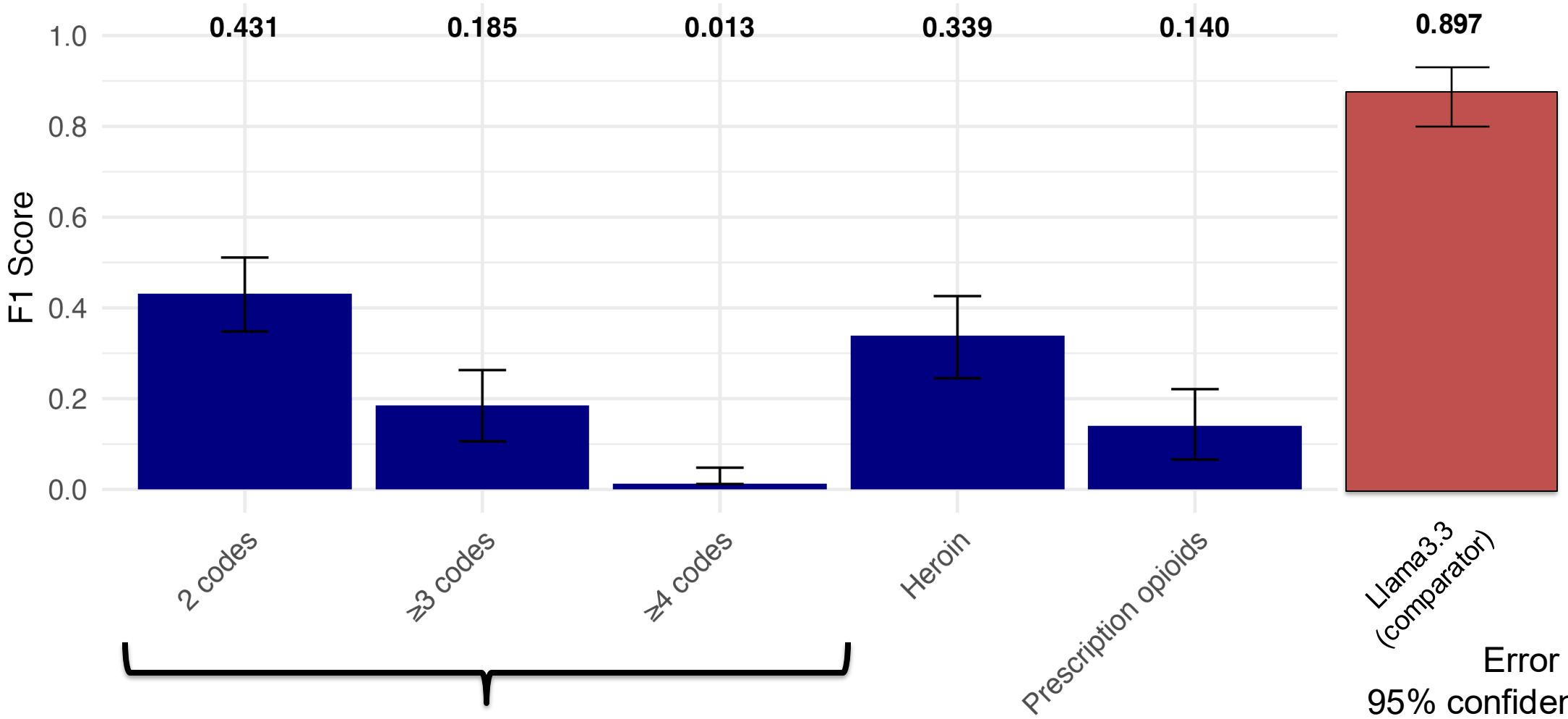
Error bars denote
95% confidence intervals

UNSW SYDNEY

UNSW Kirby Institute

ST VINCENT'S
Better and fairer care. Always.

ICD Code Performance Poor Across The Board



Error bars denote 95% confidence intervals

UNSW SYDNEY

UNSW Kirby Institute

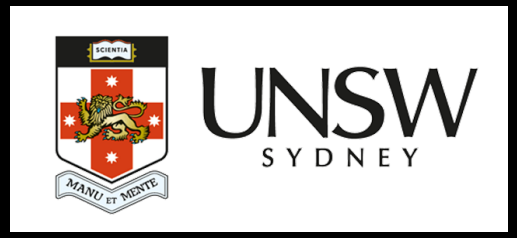
ST VINCENT'S Better and fairer care. Always.

Implications & Next Steps

- LLMs far outperformed ICD codes in identifying people who inject drugs.
- Support improved surveillance and clinical research.
- Potential for real-time case finding and targeted interventions.
- LLMs can make invisible populations visible within health systems -> a step toward equitable care.

UNSW SYDNEY
UNSW Kirby Institute
ST VINCENT'S Better and fairer care. Always.

THANKS



dgoodman@kirby.unsw.edu.au