

Beyond Structured Codes: Using Real-World Electronic Medical Record Data in Italy to Characterize Immunocompromised Patients

Poster Number

B-078

Daniel Poscover^{1*} Arjun Choleyil¹ Lorenzo Chiudinelli³ Sonia Mognaschi² Mauro Bucalo² Serena Venturelli⁴

¹ Graticule Inc., Boston, MA, USA ² Biomeris, Pavia, Italy ³ Clinical Research, Development & Innovation, ASST Papa Giovanni XXIII, Bergamo, Italy ⁴ Infectious Diseases Unit, ASST Papa Giovanni XXIII, Bergamo, Italy *Presenting author

Disclosures: This study was funded by AstraZeneca. Full author disclosures will be posted on the ISPE conference website.

Introduction

Immunocompromised (IC) patients face far higher infection rates, morbidity and mortality than immunocompetent individuals, yet "IC" encompasses a clinically heterogeneous set of conditions: hematological malignancy, solid tumor, solid organ transplantation, chronic immunosuppression, and primary immunodeficiency, each with a distinct risk profile. In Italy, however, these populations have rarely been characterized at scale from real-world EMR data, partly because structured diagnosis codes under-record chronic conditions, leaving a critical evidence gap for accurate burden estimation, risk stratification, and infection surveillance.

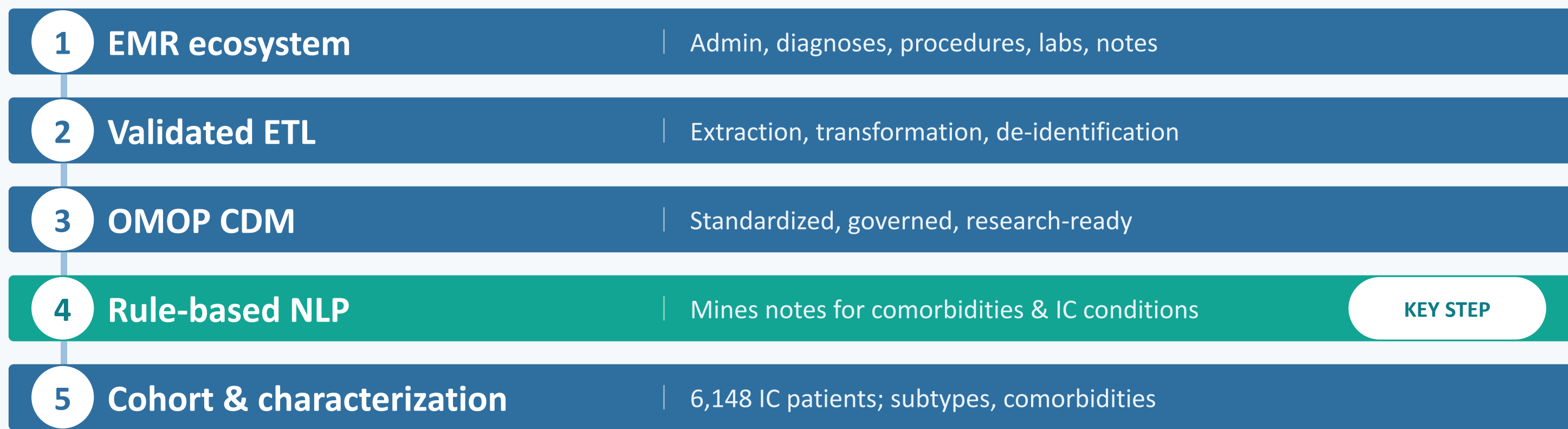
Objective

Real-world EMR data have been difficult to access in Italy, and diagnosis codes often miss chronic conditions. We aimed to enable EMR data to characterize IC patients at ASST Papa Giovanni XXIII by applying rule-based NLP to include documented IC-related facts that structured, OMOP-transformed codes would miss.

Primary: characterize the IC cohort (demographics, subtypes and 5-year comorbidity burden) from OMOP-standardized EMR enriched with NLP.

Secondary: demonstrate the value of NLP for comorbidity ascertainment where codes are incomplete.

Chart 1. Clinical Data Flow



Methods and Materials

EMR data from ASST Papa Giovanni XXIII (Bergamo), with Biomeris, were integrated into the OMOP CDM via validated ETL in a governed, secure environment. Because structured codes under-captured chronic conditions, a rule-based NLP pipeline extracted comorbidities and IC conditions from free-text notes. Patients aged ≥ 12 yrs were indexed at first visit (Apr 2022 – Mar 2023) with prior/current IC documentation. Comorbidities were assessed over 5 years and utilization over 1 year before index; respiratory-virus testing (influenza, RSV, hMPV) used multiplex panels, not sequential single-virus testing. All analyses were descriptive.

Study Parameters

Setting ASST Papa Giovanni XXIII (Bergamo); single-center, retrospective EMR study

Standardization OMOP CDM (hospital–Biomeris ETL); rule-based NLP of notes

Eligibility IC patients ≥ 12 yrs; index Apr 2022 – Mar 2023

Look-back comorbidities 5 yrs · utilization 1 yr

Data domains demographics, diagnoses, procedures, labs, prescriptions, notes

Analysis descriptive: N (%), mean (SD), median [IQR], Charlson index

Results

Of 6,148 IC patients, most had a malignancy, either hematological (49%) or solid tumor (36%), with smaller groups: organ transplant (14%), immunosuppressive therapy (5%), CAR-T/HSCT (4%), primary immunodeficiency (<1%). The cohort was older (median 66 yrs; 54% ≥ 65) and male-predominant (55%), with high comorbidity burden (17.5% Charlson ≥ 5 ; 10% with ≥ 3 distinct comorbidities). Respiratory-virus testing was infrequent (~5%); among those tested, positivity was **8.5% (influenza), 6.6% (RSV), 5.3% (hMPV)**; cohort-wide rates <0.4%.

Figure 1. Cohort Characteristics

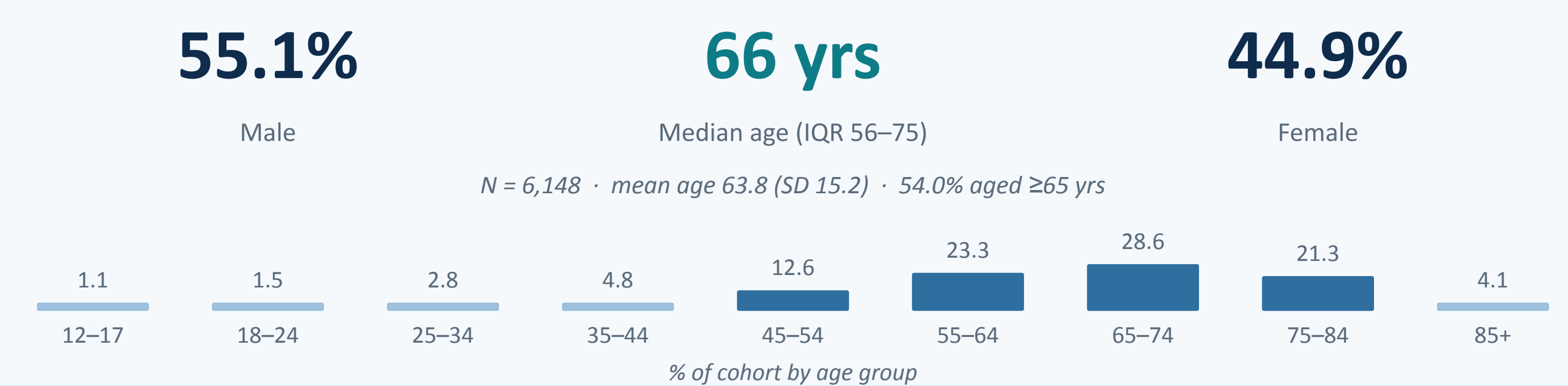


Figure 2. Immunocompromising conditions*

49%	Hematological malignancy	36%	Solid tumor
14%	Organ transplant	5%	Immunosuppressive therapy
4%	CAR-T or HSCT	<1%	Primary immunodeficiency

Figure 3. Charlson Comorbidity Index (5-yr look-back)

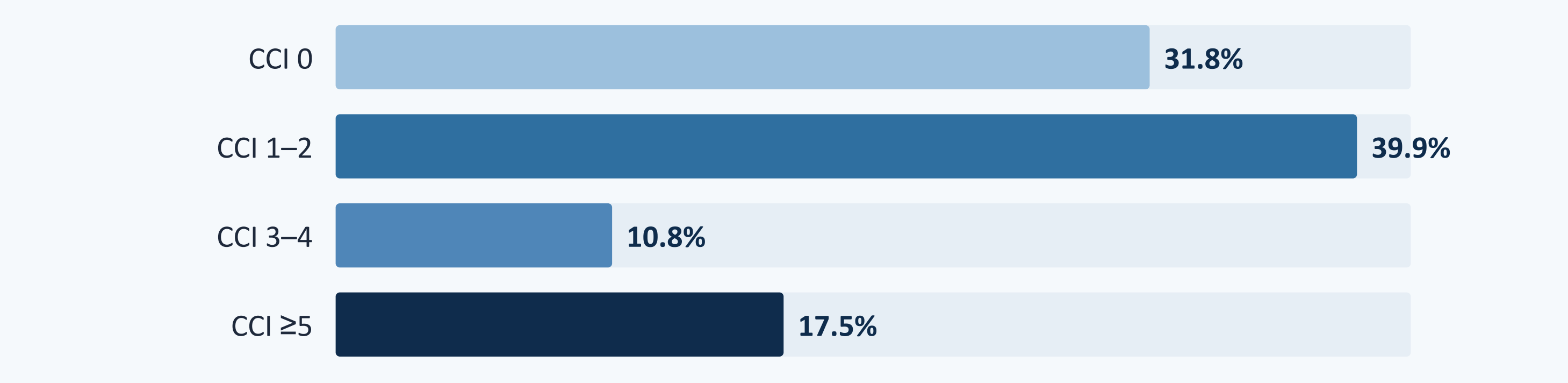


Figure 4. Comorbidity capture: codes vs. NLP*

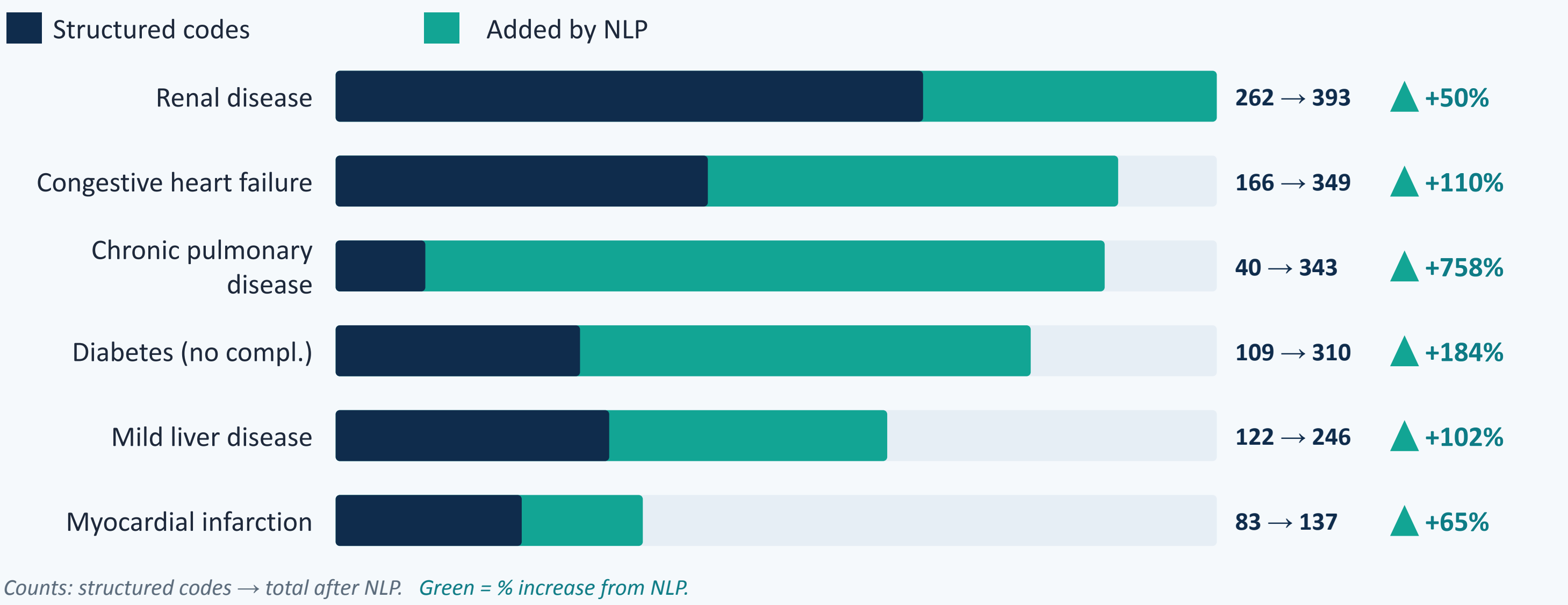


Chart 2. Respiratory-Virus Testing & Positivity

Virus	Tested (% cohort)	Positive (% tested)	Positive (% cohort)
Influenza	4.4%	8.5%	0.37%
RSV	4.4%	6.6%	0.29%
hMPV	5.2%	5.3	0.28%

Only ~1 in 20 patients was tested, yet 5–9% of those tested were positive. Influenza 23/271 · RSV 18/271 · hMPV 17/322. Influenza and RSV are tested together; hMPV is sometimes assessed separately.

Figure 5. Healthcare utilization (1-yr look-back)

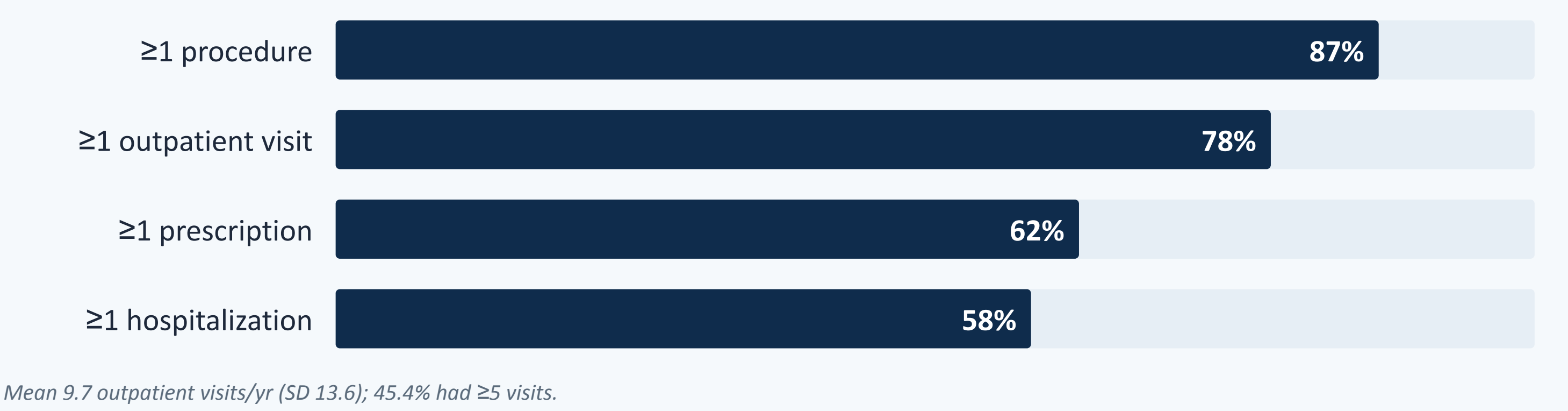
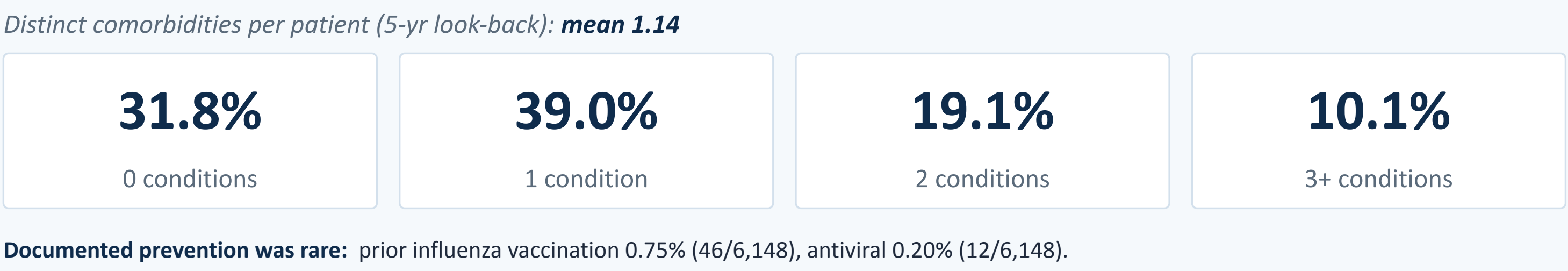


Chart 3. Comorbidity burden & prevention



Discussion

Standardizing a single hospital's EMR data to the OMOP CDM produced a research-ready resource characterizing a large, heterogeneous IC cohort at scale, but required information held in unstructured data to generate reliable inputs for analyses of IC patients. The key finding: structured codes alone substantially under-captured chronic comorbidities, e.g., chronic pulmonary disease rose +758% with NLP (40→343). Rule-based NLP of free-text notes was essential; without it, burden would have been markedly underestimated. The cohort was older, comorbid and high-utilizing. Limitations: single-center, with reliance on local narrative formats and documentation processes that constrains generalizability.

Conclusions

Standardizing EMR to the OMOP CDM with rule-based NLP enabled rigorous large-scale characterization of a complex IC population, where codes alone would understate comorbidity burden. NLP was critical for accurate ascertainment and surfaced a respiratory-virus surveillance and prevention gap warranting study.

Transferability & scalability: built on the OMOP Common Data Model with rule-based NLP over clinical notes, the workflow transfers to other OMOP-mapped hospitals and enables federated, multi-center analysis without sharing patient-level data. It offers a reusable template for scaling real-world evidence: future studies can deploy it across institutions to assemble larger, more representative immunocompromised cohorts and confirm the NLP comorbidity gains at scale, provided local notes are accessible and the NLP is tuned to each site's language and coding.

Takeaway: A transferable EMR → OMOP → NLP framework that makes hospital data research-ready, revealing what diagnosis codes alone would miss.

Contact

Daniel Poscover · Graticule Inc., Boston, MA, USA · info@graticule.life

Acknowledgements

The authors thank ASST Papa Giovanni XXIII for data access, and Biomeris for EMR standardization to the OMOP CDM.

References

1. Charlson ME, et al. J Chronic Dis. 1987;40(5):373–383. 2. Hripcsak G, et al. Stud Health Technol Inform. 2015;216:574–578. 3. Overhage JM, et al. JAMIA 2012;19(1):54–60.

* Categories are not mutually exclusive; patients may have >1 condition.