Strategic Public Policy Research Funding Scheme

策略性公共政策研究資助計劃

Project Number : 項目編號：	S2019.A4.015.19S
Project Title : 項目名稱：	Profiling the Primary, Secondary and Tertiary Prevention Needs of the Populations of Older Adults and Young-olds in Kwai Tsing and Sham Shui Po to Optimize the Operation at District Health Centers: Data-driven Service Model and Policy Design 評估葵青區和深水埗區長者及中年人的三層基層醫療預防的需要及優化地區康健中心之服務機制：數據化的服務模式及政策設計
Principal Investigator : 首席研究員：	Professor Albert LEE 李大拔教授
Institution/Think Tank : 院校 ／智庫：	The Chinese University of Hong Kong 香港中文大學
Project Duration (Month): 推行期 (月) ：	42
Funding (HK$) : 總金額 (HK$)：	4,953,002.00

Project: S2019.A4.015.19S

Profiling the primary, secondary and tertiary prevention needs of the populations of older adults and young-olds in Kwai Tsing and Sham Shui Po to optimize the operation at District Health Centers: Data-driven Service Model and Policy Design

評估葵青區和深水埗區長者及中年人的三層基層醫療預防的需要及優化地區康健中心之服務機制：數據化的服務模式及政策設計

EXECUTIVE SUMMARY

Achieving decision-centric precision public health is essential to ensure the quality of lives of our residents and the sustainability of our health and care system and addressing the needs of our aging population growing in intensity and diversity,. To this effect, a data-driven profiling of care needs and optimizing the corresponding medical and social services are essential to bringing precision to public health decision-making. Critical to bringing precision to public health is the ability to identify the heterogeneity inherent in the population's needs for disease prevention, health promotion, and health disparities reduction. However, few empirical studies have segmented the populations to optimally distinguishing one constituting sub-populations from another and sufficiently to reflect the heterogeneity critical to public health policies and interventions. In this study, the unsupervised and supervised machine learning algorithms were applied in tandem to partition the studied inpatient populations into optimal numbers of segments to capture the population's heterogeneity in clinical and acute care utilization pattern. Artificial intelligence algorithms were constructed to represent the socioecology of Kwai Tsing (K&T) and Sham Shui Po's (SSP) residents and with multi-source data analyses, the researchers of this study have identified profiles for potential primary, secondary and tertiary prevention targets of the studied populations in terms of the residents' external and internal built environment, trajectories of disease development, and their psychosocial and medical statuses. A novel combination of artificial intelligence and optimization programming has demonstrated that a data-driven optimization of medical and social services can alleviate the burden on our health and care system.

The study has identified higher 28-day re-hospitalizstion rate and a shorter median time between discharge and re-hospitalisation for those with no post-acute services. Follow up appointment at specialist out-patient clinics has shown to minimise re-hospitalisation and other types of ambulatory care services have also been found to be significiant in reducing re-hospitalisation but not very high. This would be due to different services operated in silos rather than operating as an integrated primary healthcare team. Metabolic indicators such as central obesity, lipid level, glycaemic control and co-existing chronic illnesses may also serve as markers for cancers development (liver, colorectal, gastric, kidney, pancreas). Comorbidities and prior medical service utilisations have been found to be more significant than lung functions and clinical acuity to affect the mortality trajectories of Chronic Obstructive Pulmonary Disease (COPD) patients.

Features representing internal built environment of working poor, such as the size of the living space, air quality, access to light, architectural design conducive for social connection, and age of the building, were found to assign greater statistical importance than other more commonly examined predisposing factors such as age, occupation, the severity and locations of pain, BMI, serum blood sugar, and blood pressure, for pain interference. A supervised learning-abbreviation of the Comprehensive Geriatric Assessment (CGA) can be used for screening frequent hospitalisation risk. Instrumental homecare, on its own or combined with either one or both of the other home care services, has been shown to yield the greatest costs savings compared to other services. When expressed under a joint medical-social budget perspectives, instrumental home care would reduce medical costs of HK$34.53 and HK$85.03 for every dollar invested in instrumental-restorative home care respectively.

A deep learning model with multiple input channels to capture the hierarchical relationships among the demographical, medical, behavioural, psychosocial, and built environment levels of an individual's socioecology

was developed and found that built environment features were assigned greater statistical importance to general health compared to features associated with one's socio-demographics, health, and health-related behaviours and service utilization, and were also found to link to COVID-19 case count. The built environment that put residents at risk for poor general health and also put them at risk for COVID-19.

Modifying the risk factors putting population at risk not only decreases the risk of developing respective diseases and also preventing deterioration of other co-exisitng diseases, e.g., adverse living conditions related to pain among working poor and surge of Emergency Department (ED) waiting time. Screening out the population at risk for developing a particular disease would also have impact on co-existing health conditions or development of other diseases such as metabolic indicators and cancers development. Control of co-morbid conditions can have impact of life expectancy of other conditions such as COPD, and increased risk of developing other diseases such as cancers. Profiling the primary, secondary and tertiary prevention needs would guide the development of primary healthcare meeting the needs of the population by integrating preventive, promotive, curative, rehabilitative, and palliative healthcare services.

Most notably, in addition to the traditional means of disseminating the findings to our fellow academics, the general public, key community leaders, and governmental decision-makers, a predictive A.I. platform has been developed using novel algorithms to integrate the findings across different aspects of the socioecology of the studied districts to aid clinical and policy decision making for the district offices of our partnering districts. The researchers had exchange with government officials, opinion leaders and residents of the studied communities on analyses of different databases and knowledge leading to observations:

- While the profiles of those who “fell through the cracks” of our medical and social systems are consistent over time and across districts, the development of policies to address them should be district-specific to ensure the successes of their implementation.
- Currently, clear service gaps exist for patients aged 50+, young old with re-hospitalisation soon after discharge without post-acute care or ambulatory care services in silo.
- Community-dwelling elderly who lacks independence, and those whose quality of life were interfered by chronic pain.
- Targeted screening and optimided pairing between service and care needs can result in enormous cost savings
- Metabolic dysfunction and smoking (includes second-hand smoking) should be targeted for primary and secondary prevention across all districts in HK.
- This learning algorithms revealed that not only did metabolic dysfunction and smoking contribute directly to the development of diabetes and other chronic illnesses, but they have also contributed indirectly but consistently to cancer of the digestive, nephrology and urology systems.
- Government should invest in data-driven operations research and analytics to enable intelligence-assisted service assignment and bring precision to preventive care. We also recommend.
- The primary and secondary prevention should increase in precision by targeting specific housing estates in the order of their residents' potential primary and secondary prevention needs, as our socio-ecologically anthropomorphided deep learning algorithms have demonstrated that aspects of the residents' internal and external environment may put residents at an elevated risk for poor general health, and COVID-19 infection (which has become the most imminent primary prevention needs of Hong Kong residents since the current project has commenced).

隨著我們老年人口需求的增長和多樣化，實現以決策為中心的精確公共衛生至關重要，以確保我們居民的生活質量和我們的健康和護理系統的可持續性。為此，對護理需求進行數據驅動的分析，並優化相應的醫療和社會服務，對於提高公共衛生決策的精確性至關重要。實現公共衛生的精確性的關鍵在於能夠識別出人口對疾病預防、健康促進和減少健康差距的需求中的異質性。然而，很少有實證研究將人口劃分為最優的子群體，以便最大程度地區分一個構成的子群體和另一個，並充分反映出對公共衛生政策和干預至關重要的異質性。無監督和監督的機器學習算法被並行應用，將研究的住院人口劃分為最優數量的片段，以捕捉人口在臨床和急性護理利用模式中的異質性。人工智能算法被構建來代表葵青（K&T）和深水埗（SSP）居民的社會生態，通過多源數據分析，本研究的研究者已經確定了研究人口的潛在一級、二級和三級預防目標的概況，包括居民的外部和內部建築環境、疾病發展的軌跡，以及他們的心理社會和醫療狀況。人工智能和優化編程的新穎組合已經證明，數據驅動的醫療和社會服務的優化可以減輕我們的健康和護理系統的負擔。

該研究發現，那些沒有接受急性期後服務的人的 28 天再住院率較高，出院後再住院的中位時間較短。專科門診的隨訪預約已經顯示出可以最小化再住院，其他類型的門診護理服務也被發現在減少再住院方面具有重要意義，但並不是很高。這可能是由於不同的服務在孤立的情況下運作，而不是作為一個整體的初級衛生保健團隊運作。中心性肥胖、血脂水平、血糖控制和共存的慢性疾病等代謝指標也可能作為癌症發展（肝、結腸、胃、腎、胰）的標誌。已發現合併症和先前的醫療服務利用比肺功能和臨床急性症狀更能影響慢性阻塞性肺病（COPD）患者的死亡軌跡。

代表工薪階層內部建築環境的特徵，如居住空間的大小、空氣質量、光線的獲取、有利於社交的建築設計和建築的年齡，被發現比其他更常被檢查的疼痛干擾的易感因素賦予了更大的統計重要性，如年齡、職業、疼痛的嚴重程度和位置、BMI、血清血糖和血壓。全面老年評估（CGA）的監督學習縮寫可以用於篩查頻繁住院的風險。儀器家庭護理，無論是單獨的，還是與其他一種或兩種家庭護理服務結合，都被證明與其他服務相比，可以產生最大的成本節約。當以聯合醫療-社會預算的角度表達時，儀器家庭護理可以減少每投資一元在儀器恢復家庭護理上的醫療費用 34.53 港元和 85.03 港元。 開發了一個深度學習模型，具有多個輸入通道，可以捕捉個體社會生態的人口統計、醫療、行為、心理社會和建築環境層次之間的層次關係，發現建築環境特徵被賦予了比與個人的社會人口統計、健康和健康相關行為和服務利用相關的特徵更大的統計重要性，對一般健康，並且也被發現與 COVID-19 病例數有關。將居民置於健康狀況不佳的風險的建築環境也將他們置於 COVID-19 的風險之中。 修改將人口置於風險的風險因素不僅可以降低發展相應疾病的風險，還可以防止其他共存疾病的惡化，例如，與工薪階層疼痛相關的不良生活條件和急診等待時間的激增。篩查出有發展某種疾病風險的人口也會對共存的健康狀況或發展其他疾病產生影響，如代謝指標和癌症的發展。控制合併症可以影響其他疾病的預期壽命，如 COPD，並增加發展其他疾病的風險，如癌症。對一級、二級和三級預防需求的分析將指導初級衛生保健的發展，以滿足人口的需求，通過整合預防、促進、治療。

LAYMAN SUMMARY ON POLICY IMPLICATION AND RECOMMENDATION

The medical and social care costs of the K&T and SSP districts can be greatly reduced by improving the precision and efficiency of surveillance and community-based interventions on 1) the signs and complications of metabolic dysfunctions and 2) the lack of independence in instrumental activities of daily living. We recommend that HK Government to work with local District Offices to make available support that targets the surveillance and community-based interventions for metabolic dysfunctions and functional dependence, especially for those 50+ who suffer from COPD, diabetes, or cancers of the digestive, nephrology and urology systems for both districts, and chronic pain and mental health for K&T residents. Specifically, the precision and efficiency of surveillance and community-based interventions can be enhanced by artificial intelligence systems and learning algorithms created with the support of the current study. In addition, community-based interventions can be achieved via the

training of non-professional volunteers and providing them with access to easy-to-use artificial intelligence system - for example, the project team is currently supporting the District Offices of K&T and SSP to enable their respective Care Team to perform surveillance.

透過提高對於 1) 代謝功能障礙的表徵和併發症 和 2) 缺乏獨立自理能力 的監測和社區介入措施的精確度和效率，葵青及深水埗區 的醫療和社會照護成本可以大大降低。

我們建議香港政府與 各區民政事務處合作，提供針對代謝功能障礙和缺乏獨立自理能力的監測和社區幹預的支持，特別是針對那些 50 歲以上患有慢性阻塞性肺病、糖尿病，或消化，腎臟和泌尿科系統癌症。除上述病人以外，葵青區比起深水埗區及香港各區更加需要對於長期痛症和心理健康有更多的支援。

具體來說，在目前研究的支持下創建的人工智慧系統和學習演算法可以提高監視和基於社區的干預措施的精度和效率。 此外，以社區為基礎的干預措施可以透過培訓非專業志工並為他們提供易於使用的人工智慧系統來實現——例如，計畫團隊目前正在支持葵青及深水埗的民政事務處，以使各自的護理團隊能夠進行監控

The report will be presented as different chapters.

Chapter 1 reviews the background of the study and review of current literature.

Chapter 2 is preamble of the study project.

Chapter 3 describes the methodology.

Chapter 4 presents the findings under different sections covering the key objectives.

Chapter 5 discusses the findings and policy recommendations.

Chapter 6 summaries the knowledge transfer and dissemination.

Chapter 7 is the conclusion.

Principal Investigator: Professor Albert Lee

Co-Investigators: Professor Eman Leung**,** Professor Hector WH Tsang

Other co-investigators at the time of application: Professors Stella WC Cheng, Stephen WK Chiu, CB Chow, Diana TF Lee

Papers published/To be published

Leung E, Lee A, Tsang H, Wong MC (2023). Data-driven service model to profile healthcare needs and optimise the operation of community-based care: a multi-source data analysis using predictive artificial intelligence. *Hong Kong Med J*, 29(6):484-6. Available from: https://doi.org/10.12809/hkmj235154

Leung E, Lee A, Liu Y, et al (2024). Impact of environment on pain among the working poor: Making use of random forest-based stratification tool to study the socioecology of pain interference. *International Journal of Environmental Research and Public Health* 2024 *21*, 179. https://doi.org/10.3390/ijerph21020179

Leung E, Guan J, Zhang Q, Lee A, Tsang H, Ching CC, Yee H, Yeoh EK, Chen FY (2024). Screening for Frequent Hospitalization Risk among Community-dwelling Elderly between 2016 and 2023: Machine learning-driven item selection, scoring system development, and prospective validation. *Frontier in Public Health* Forthcoming

Leung et al (2024). Data-driven optimization of a population-level allocation of post-discharge services for diverse patient segments to improve patient outcome and maximize medical cost savings: A case for decision-centric precision public health. *Frontier in Public Health* Forthcoming.

Leung et al (2024). The Valuation of Elderly Homecare Services Under a Joint Medical-Social Budget Perspective. *Frontier in Public Health*

Lin, S., Bakker, M. & Leung, E (2022). Robust meal delivery service for the elderly: a case study in Hong Kong. *Sci Rep* **12**, 21546. https://doi.org/10.1038/s41598-022-25924-6

Yau S, Leung,E, Wong MCS, et al (2024b) Differential metabolic dysfunction profiles and site-specific risk of obesity-related cancers in patients with diabetes" *BMJ Open* Apr 3; 14(4):e082414. doi: 10.1136/bmjopen-2023-082414.

Preprints avaiable

Guan, Jingjing & Leung, Eman & Kwok, Kin-On & Hung, Chi & Lee, Albert & Chong, Ka & Yam, Carrie & Cheung, Clement & Tieben, Hendrik & Tsang, Hector & Yeoh, EK. (2023). Quantifying The Risk of General Health and Early Covid-19 Spread in Residential Buildings With Deep Learning and Expert-Augmented Machine Learning. *10.2139/ssrn.4432448*

Leung et al (2024g) on Article on "Deep Learning Approach to Forcasting COVID-19 Cases in Residential Buildings' of Hong Kong: A Multi-soueces Data Analysis of the Differential Roles of Environments and Sociodemographic during the Emergence and Resurgence of the Pandemic. *ArVix* https://arvix.org/abs/2403.15759

Leung E, Guan JJ, Kwok KO, Hung CT, Chng CC, Tsang H, Yoeh EK, Lee A. (2024g) Article "Analysing the Variations in Emergency Department Boarding and Testing the Transferability of Forecasting Models across COVID-19 pandemic Waves in Hong Kong: Hybrid CNN-LSTM approach to quantitfying building-level sociological risk. *ArXiv: 2403.13842(cs)* (http://arvxiv/abs/2403.13842).

Invited presentation

Lee A. *Healthy City Movement for Sustainable Urban Movement* A Two Part International Series Planning Healthy City Seminar, Organised by the Hong Kong Institute of Planners 27 February 2024

CHAPTER 1
BACKGROUND, INTRODUCTION & REVIEW OF LITERATURE

The emerging interest in bringing precision to public health notwithstanding, empirical studies on the topic is scarce. The concept was first introduced into the academic literature and Centre for Disease Control's (CDC) lexicon in 2015 by Khoury [2016]. While the emerging body of data-driven precision public health research shares the objective of bringing precision to the prevention of diseases, promotion of health and the reduction of health disparities in populations," the lack of a robust data-driven approach to segment the population into homogeneous sub-population has always been considered the root of the evidentiary challenges faced by precision prevention, e.g., it has been noted that since prevention guidelines are typically designed for average individuals in the population, developing evidence-based recommendations targeting specific sub-population defined by traditional risk factors is challenging when data on the balance of benefits and harms is lacking. Studies on sub-populations were usually based on relatively small number of research participants. Consequently, stratification of disease risk based on risk factors leaves most people either "slightly above average or "slightly below average" risk, making the corresponding evidence accumulated challenging for recommending different courses of action to preserve health for most people.

While it has never been attempted as a means to bringing precision to defining sub-populations for targeted public health interventions, patient segmentation research has the potential to achieve patient-centeredness with sophisticated analytic methodology that optimizes the level of granularity in defining sub-populations with parameters relevance to population health [Vuiki et al, 2016]. There were also exceptions among these segmentation studies, wherein only expert opinions and literature reviews, rather than sophisticated analytics, were deployed to segment patients [Chong et al, 2019; Low et al, 2018]. Notwithstanding its objective of sub-grouping patients according to their diverse care needs, the literature on patient segmentation fails to operationally define patient care needs with respect to patients' clinical acuity and complexity. Rather, the studies commonly segment their respective sample only with age and specific clinical conditions pre-selected by authors [Chong et al, 2019; Low et al, 2018], thus failing to distinguish high-need patients whose age and clinical conditions were the same but differed in the intensity of required care due to, for example, co-existing chronic illnesses; or only with care costs or service utilization parameters totaled across different care locale on a patient journey where they were sourced without distinguishing the variation in clinical acuity and complexity of patients serviced by different care locales [Vuik et al, 2016; Low et al, 2018; Nnoaham and Cann, 2020; Davis et al, 2023].

Decision-centric precision public health, while being essential to public health practices and health policy decision making, lacks both theoretical frameworks and empirical research which is unfolded by this study using machine learning-informed joint medical-social perspective to drive precision public health research and interventions in Hong Kong with population data [Leung et al, 2024a]. Adding to the challenges of exercising decision-centric precision public health is where there is a lack of data to drive precisions in public health planning and resource allocation. In the context of Hong Kong, the area where data-driven planning and resource allocation decision most needed is also the area that experience most scarcity in relevant data – the district-based primary care services that have recently been devolved to Non-Government Organsiations (NGOs) to operate in the form of District Health Centres (DHCs). Not only were the NGOs that operate the DHCs either not routinely collecting data for operations research and management purposes, nor, even when collected, they were not accessible because of a non-digital nature; and the NGOs also fail to routinely collect all relevant aspects of DHCs' clients and spheres of the district's socioecology that impact on DHCs' clients or potential clients.

This is particularly of note as CDC's socio-ecological framework for prevention has highlighted the importance of the person-level and other spheres across the person's socioecology to different tiers of prevention. This study has two papers forthcoming using this kind of socio-ecological framework for prevention (can refer to those

papers for details) [Guan et al, 2024; Leung et al, 2024b]. The challenges of collecting data with sufficient volume and quality for bringing precision to DHC's planning and resource allocation have been further exacerbated by the emergence of the COVID-19 pandemic, whose peaks of outbreaks coincide with the opening of the first two DHCs at Kwai Tsing and SSP. Resulting in the cancellation of routine services, poor attendance when services re-open, and ad hoc service provisions.

Hence, with a novel application of machine learning and deep learning methodologies, we aimed to bring precision to the planning and operations of DHCs by profiling prevention needs of the primary (when no clear indication that the disease process has begun), secondary (disease process has begun but not yet requiring hospitalization, and tertiary (patients deteriorate further, resulting in re-hospitalization and/or complications) nature in the regional populations of older adults (65+) and young olds (50-64). Specifically, the current project proposed to profile the prevention needs of the studied populations with respect to different spheres in CDC's socio-ecological framework for prevention, which include: the person-level, the living environment and the socio-demographics of their community, the service ecology, and the relevant government policy.

In addition, the multi-dimensional profiles developed from our multi-source data analysis (which triangulated analyses performed across databases ascertained from different socioecological sphere of the studied districts) were then served as the basis for developing data-driven optimization programming to match between residents' prevention needs on the one hand, and service types and quantity on the other hand. This study provides a review on this emerging area of applying operations research and management sciences to health care, and the current research gap. Two studies have been conducted that our project is trying to contribute to [Lin et al, 2022; Leung et al, 2023b].

The study has been approved by Kowloon West Research Ethics Committee of Hospital Authority, Hong Kong (Ref: KW/EX-22-030(170-06))

Objectives

To achieve decision-centric precision public health, four objectives were proposed:

Objective #1: To develop clinical and functional profiles of Kwai Tsing (K&T) and Sham Shui Po (SSP) residents by conducting repeated assessments of a cohort of 1026 individuals from the K&T & SSP Districts

Objective #2: To apply an artificial intelligence algorithm to map the trajectory of how primary, secondary and tertiary prevention needs evolve in the studied districts in order to forecast the growing demands of medical and social services of the population

Objective #3: To make available algorithms that can optimize the cost-effectiveness of the service mix offered at the District Health Centers of K&T and SSP, which may enable the operators of K&T and SSP District Health Centers to better integration between medical and social services of K&T and SSP

Objective #4: To establish a novel methodological framework and a technological platform for profiling primary, secondary and tertiary prevention needs using artificial intelligence for targeted needs assessment and risk identification, and data-driven policy design for community-oriented primary care

Changes from the original protocol resulting from COVID-19 and their impacts

In summary, we expanded the project scope to include not only the K&T residents as initially proposed, but also populations from the SSP district. This expansion was in response to requests from our stakeholders, including Dr. CB Law of the Kowloon West Cluster and Dr. Fan Ning, the Clinical Director of SSP DHC. The broadening

of the scope to SSP led to the development of a strong working relationship and an advisory role for the District Officer of SSP, culminating in the commissioning of a predictive AI platform by the SSP district Office. This platform translates the academic output springboarded by SPPR into assistive intelligence for policy and service deployment decisions. And as it turns out, the SSP-commissioned platform also enables us to benefit K&T despite the different needs of the two districts and the challenges posed by COVID-19, personnel changes, and the re-evaluation of commitments, our AI platform is now being made available to K&T DHC through working with its Chairman, Mr. Chow Yick-hay, and to SSP DHC as brokered by SSP District Officer. But more directly impacting the residents of the studied districts, we are working with Mr. Chow Yick-hay and SSP District Officers to enable the Care Team that they lead in their respective districts to adopt the AI platform built from studies funded by the SPPR to address the service gaps of each DHCs, as identified using the tertiary-wide and cluster-based EHR records as well as assessments performed by our partnering NGOs. None of these knowledge translation activities would be possible if not were for the support of SPPR, both in terms of funding and the agile re-tooling of our research strategies.

We also broadened the study scope in response to the long-term presence of the virus in the city and the potential changes in preventive care priorities and clinical case mixes. We expanded our research framework to better understand the socioecology of the studied population and included risk to COVID-19 infection and hospital utilization forecasting associated with COVID-19 in the study, as part of the profiling of studied districts' preventive care needs. Because of the socioecological framework we adopted, we have aligned our methodology using multi-sourced data analysis and multi-headed hierarchical CNN, which we have anthropomorphized, explainable, and expert-augmented. All of these are innovation that made possible only by SPPR support.

In addition to the change in scope, we adapted to COVID-19 and the corresponding public health and social measures. These adaptations included giving up direct recruitment of discharged patients from KWC hospitals, engaging new social service partners, developing natural language processing algorithm to extract meaningful information from unstructured clinical notes in cases when instruments in our assessment package has gone beyond the NGOs' service scope, and using multi-sourced data analysis on profiles developed from different population databases with a socioecological framework-anthropomorphized AI to overcome the biases of sampling from social services resulting from COVID-19.

Consequently, while COVID-19 and the corresponding public health and social measures have inspired innovations, the SPPR project fund has not been spent on generating any new proprietary project data. Instead, SPPR's support made possible the secondary analyses and triangulation of data extracted from the public domains to develop analytic entities to guide the analysis of patient and client records that Hospital Authority and our partnering NGOs collected as part of their standard care. As a result, no records from the NGOs or Hospital Authority that the SPPR has funded to analyze could be made available to CEPU or the public due to Hong Kong's Personal Data (Privacy) Ordinance (PDPO) and their internal policies. The partnering NGOs and Hospital Authority can only consent to letting us analyze and report the findings resulting from data under their custodianship as long as the strict confidentiality agreements were followed that safeguard the anonymity of their clients (who had provided consent to the NGOs and Hospital Authority to have their data analyzed and reported only in an anonymized and aggregated format for the purposes of service improvement, education and research). However, because of the partnership that SPPR has supported, the two NGOs had upgraded their routine record-keeping and assessments, making the clientele of the two NGOs direct beneficiaries of SPPR.

CHAPTER 2 PREAMBLE TO OUR RESEARCH METHODOLOGY

To achieve this Strategic Public Policy Research (SPPR) project aim of profiling the primary, secondary and tertiary prevention needs of the populations of older adults and young-olds in Kwai Tsing and Sham Shui Po to optimise the operation at DHCs, multi-source data analyses were performed with predictive artificial intelligence algorithms with data sourced from the public domain or the partnering medical and social services' historical and prospectively collected (within the timeframe of our SPPR project) records. Hence, while the SPPR project fund had not been spent on generating any new proprietary project data, SPPR's invaluable support had made possible the secondary analyses and triangulation of data extracted from the public domains and data collected by the Hospital Authority and our partnering NGOs from their clients as per their respective routine record-keeping on service utilisation and health and social risk assessments. The active support of partnering NGOs was deeply appreciated, which had not only expanded the scopes of their routine record-keeping and assessments during the project period to align with this project's latest set of predictive and outcome features as approved by CEPU (as far as their lie within the service scope of the NGOs and permitted by their resources, otherwise, to ascertain information that lies beyond the NGO's service scope but is nonetheless relevant to the project, NGO's custodianship natural language processing (NLP) algorithms was developed and deployed to case notes (for example, classifying those who lack independence in instrumental activities of daily living from unstructured text). The NGO had also consented to the analysing and reporting of the data in their custodianship as long as the strict confidentiality agreements were followed that safeguard the anonymity of their clients (who had provided consent to the NGOs to have their data analyzed and reported only in an anonymized and aggregated format for the purposes of service improvement, education and research). For the remainder of this section, an excerpt from the editorial on a recent issue of HK Medical Journal (Leung et al, 2003, https://www.hkmj.org/abstracts/v29n6/484.htm) will be provided as an overview of the methodology we selected and motivation behind the selection, as well as the findings and impact resulted from such selection. The editorial, titled "Data-driven service model to profile healthcare needs and optimise the operation of community-based care: a multi-source data analysis using predictive artificial intelligence" was written by the project team (Leung et al, 2023b) to highlight how the support of SPPR has enabled the research team to overcome sizable contextual challenges that other health and social service researchers may face, namely the siloed medical and social services whose inter-bureau sharing of data, and the sharing of their data with academic institutions, is regulated by Hong Kong's Personal Data (Privacy) Ordinance (Cap. 486), and, of course, the burdens and restrictions on medical and social services brought by COVID-19 pandemic.

“As the needs of our ageing population grow in intensity and diversity, there is a need to achieve precision in public health via data-driven profiling of population-level preventive care, while optimising medical and social services to address those needs. These initiatives will maximise population health and minimise health care costs. Nevertheless, population-level precision public health research is rare; its application to drive service planning and deployment at the population level is even rarer. Thus, with support from the Strategic Public Policy Research Funding Scheme managed by the CEPU (previously the Policy Innovation and Co-ordination Office of the Hong Kong SAR Government, we initiated a research programme to fill the gap in precision public health research and practice by triangulating data that represent population-level socioecology, such as personal-level clinical and functional data, relational-level data for individual households, community-level data regarding socio-demographic characteristics and physical living environments, data describing organisations that meet population-level needs, and data reflecting the impacts of governmental policy. We sought to identify individuals who can receive the greatest benefit from primary, secondary, and tertiary preventive care. The resulting profiles could inform population-level planning and allocation of the three tiers of preventive care programmes.

Nevertheless, our research objectives were confronted with challenges related to the following contextual factors: (1) the inherent biases and quality of real-world data extracted from medical services’ Electronic Health Records (EHRs) and social services’ record systems; (2) the fragmentation among services (and their respective databases) which are required to address needs arising from specific aspects of population-level socioecology, including the

distinct medical and social needs that our siloed medical and social services seek to address; and (3) the coronavirus disease 2019 (COVID-19) pandemic and the associated social and public health measures which emerged shortly after project initiation and have persisted throughout its life cycle. To overcome these challenges, we adopted a multi-source analytical approach (Noi et al, 2022), whereby parallel and iterative analyses were performed across databases representing different socioecology aspects at the resident level. Specifically, an analytical profile developed in one database was applied to other databases with the goal of identifying research questions and facilitating the selection of corresponding features and analytics. The findings from multiple siloed databases could be triangulated to coherently address individual research objectives. In addition, where applicable, parameters extracted from siloed databases were integrated to model particular outcomes using our artificial intelligence (AI) algorithm, for which the input architecture was anthropomorphized (Glikson and Woolley, 2020) according to spheres described in the socioecological prevention framework of the Centers for Disease Control and Prevention. This approach enabled structuring of the hierarchically interrelated input layers.

In the following text, we describe our multi-source analytical approach and emerging findings from our research programme. Although the academic outputs of our research programme are in various stages of peer review, this description of a data-driven process to formulate research questions and develop sampling frames for examination across siloed databases in the construction of a population-level coherent care profile may serve as an alternative approach for other researchers to consider when they face similar contextual challenges in population-level precision public health research.

For example, using the study populations' EHRs (obtained via the Hospital Authority Data Collaboration Laboratory), we applied unsupervised and supervised machine learning algorithms in tandem to identify tertiary prevention needs and the service gaps that prevent those needs from being met in the study populations. Our analyses revealed that the highest re-hospitalisation rates (>80%) and the shortest times between discharge and re-hospitalisation occurred in sub-populations of patients who lacked specific ambulatory or post-acute services. Nonetheless, these services were also available to patients who shared similar clinical and utilisation profiles but exhibited significantly lower re-hospitalisation rates. Among the sub-populations with high re-hospitalisation rates and low utilisation of re-hospitalisation-mitigating post-discharge services, one had a typical profile (i.e,. population segment medoids) of patients aged 50 to 64 years with musculoskeletal pain–related disorders as primary diagnoses. These patients more frequently exhibited a history of multiple chronic illnesses and higher clinical complexity at index hospitalisation compared with other patients who had similar clinical and acute care utilisation profiles.

The profiling of sub-populations who fell through the service gaps and were rehospitalised at the highest rate enabled us to bring precision to tertiary prevention efforts and subsequently perform data-driven optimisation of population-level post-discharge service allocation, thereby minimising medical costs. Furthermore, the profile we constructed from EHRs could also be applied beyond medical settings to identify potential secondary prevention targets that may exacerbate the evolution of an underlying disease process, such that it interfered with quality of life among individuals who matched the EHR-based and machine-constructed profile, ultimately triggering health-seeking behaviour.

Thus, in a non-medical setting, we recruited residents of the study population aged 50 to 64 years who had musculoskeletal pain, according to community-based primary care clinicians. In addition to the residents' socio-demographic characteristics, behavioural health, and co-morbid chronic illness statuses, clinicians also assessed anthropometric measures and biomarkers of metabolic dysfunction that are often direct or indirect precursors to the most common forms of chronic illnesses. These factors were included as predictive features in a random forest model for selection and risk-scoring of potential secondary prevention targets that could mitigate the exacerbation of pain symptoms. The model also included features representing various aspects of the residents' living environments, which were separately parameterised and initially selected by our AI algorithm according to the following constraints: (1) they were sourced from multiple public domain datasets that belonged to

governmental agencies such as the Census and Statistics Department, Housing Authority, Lands Department, Department of Health, and District Offices; (2) they were organised as layered input into a multi-headed hierarchical convolutional neural network, with an anthropomorphised architecture that captured the study population's internal and external built environments and socio-demographic profiles; and (3) they were selected according to the statistical importances of their unique and combined contributions to residential building-level aggregates of general health based on census data and COVID-19 case counts from the Department of Health.

Finally, after parameterisation and selection in accordance with their degrees of importance to the population's general health and COVID-19 susceptibility, features representing the built environments of the study district's residential buildings were processed as follows: (1) they were entered into a random forest model together with the aforementioned individual-level measures to compare their respective importances in the onset of pain interference; and (2) they were scored according to their individual and combined adverse health effects, then assigned to individual residential buildings in the study district for optimised allocation of local primary prevention programmes.

Our analyses revealed that, although features representing residents' socio-demographic characteristics and metabolic dysfunction had high importance with respect to the presence of pain interference in various residential quality of life domains, their feature importances were secondary to the importances of built-environment features, such as living area size, air quality, access to light, architecture conducive to social connectivity, and building age. In addition to scoring the risk of pain interference for individual residents, we scored the built environment of each building in public housing estates within the study district according to the likelihood that its residents would experience sufficient pain to interfere with their quality of life. This scoring approach can inform service planning in geospatially targeted secondary pain prevention programmes.

Patients with chronic obstructive pulmonary disease who exhibited high clinical complexity and multiple co-morbidities were another sub-population who typically exhibited high rehospitalisation rates and low utilisation of re-hospitalisation-mitigating post-discharge services. This patient profile was used to guide the recruitment of study district residents outside of medical settings, enabling examination of the evolution of disease processes and hospitalisation trends among asymptomatic and symptomatic community residents. Together with the findings regarding musculoskeletal pain and health-related effects of the built environment, our work has provided the basis for a predictive AI platform that was commissioned by the Sham Shui Po District Office to support its social health surveillance and policy decision needs. Additionally, our work has been incorporated into an algorithm deployed at community diagnosis events hosted by the Sham Shui Po District Office and at events co-hosted by the Kwai Tsing Safe Community and Healthy City Association and the Kwai Tsing District Office."

CHAPTER 3: RESEARCH METHODOLOGY

The research methodology was designed to achieve the above research objectives within the context of 1) the socioecology of the studied districts' residents, 2) the real-word data extracted from medical services' EHRs and social services' system, 3) the fragmented medical and social services and their siloed databases that constitute HK healthcare system, and, most of all, 4) the COVID-19 pandemic and the associated social and public health measures emerged shortly after the commencement of the project and have lasted throughout the project life-cycle. Specifically, the above research context created barriers to collecting data from medical and social services, assuring data quality, and integrating data across services and departments, which could potentially post challenges to our objective of profiling prevention needs across different aspects of a person's socioecology. Hence, to mitigate the potential challenges, multi-source data analytic approach was applied to triangulate data and analytic profiles developed from the data sourced from databases representing different aspects of the residents' socioecology:
(1) data collected from clinical, psychosocial, functional, quality-of-life assessments performed on 508 (56% aged 50-64; 44% 65+) and 533 (38% aged 50-64; 62% 65+) clients from K&T and SSP's community services, respectively (in total, data from 3,123 assessments were available for analysis);
(2) medical records of the inpatient populations whose home addresses were in the studied districts (made accessible by Hospital Authority Data Collaboration Lab; HADCL);
(3) client records extracted from the operational databases of partnering community services; and
(4) public domain data belonging to different governmental agencies (Census and Statistics Department, Housing Authority, Lands Department, Department of Health, and District Offices).

Specifically, with parallel and iterative analyses performed across databases that represented different aspects of the residents' socioecology, we applied the analytic profiles of the residents developed from one database to drive the identification of research questions and the corresponding selection of features and analytic models in other databases. Consequently, the analytic profiles constructed from different databases (representing elements from different spheres of the residents' socioecology) served as the "entities" whose matching across these databases enabled multi-source data analysis to be performed and research questions emerged from each data source be converged under the four project objectives for their achievement.

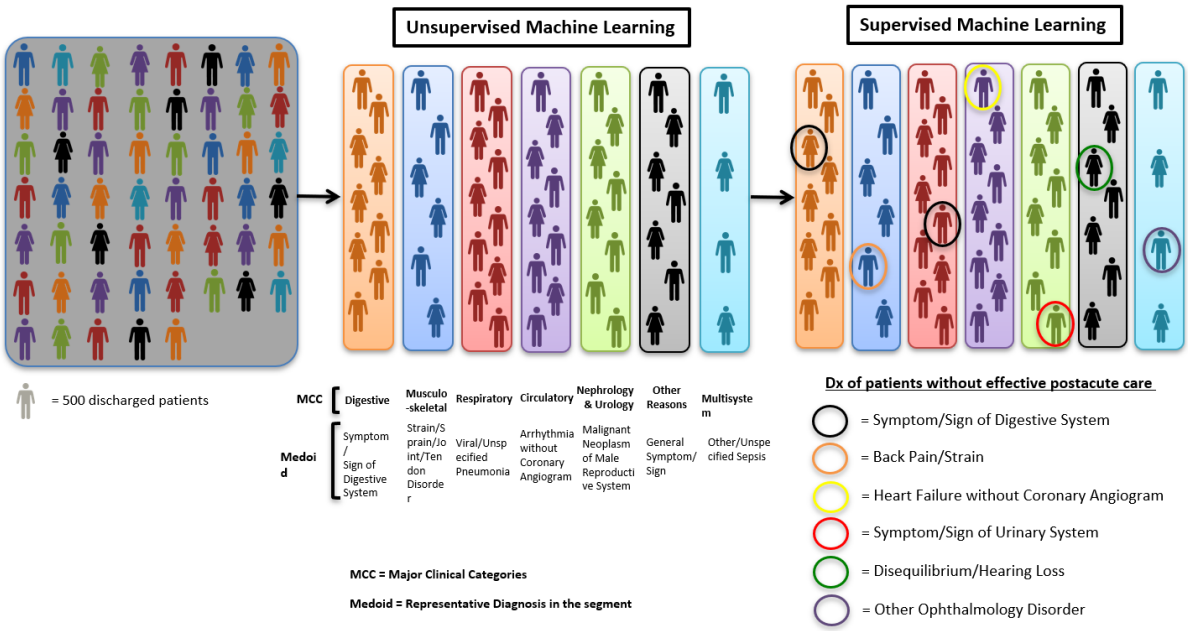
The following provides an overview of the method adopted in the individual studies conducted to achieve the project objectives within the context of the studied districts' socioecology.

3A. Methods of Studies Motivated by Objective 1

The following is the method adopted by the study reported in the article titled "**Driving precision public health research and intervention in Hong Kong with population data: A machine learning-informed joint medical-social perspectives,"** which has been submitted to *Hong Kong Medical Journal* and is currently under revision after review (Leung et al, 2024a,).

Methodological overviews

In the current study, a combination of unsupervised and supervised learning algorithms were deployed (please refer to the figure below for a schematic representation of the methodology used herein).



First, unsupervised learning algorithms were applied to the studied population to identify their typical patients using a comprehensive set of clinical and *acute-care* utilization parameters, whereby patients similar to the typical patients in terms of the set of clinical and acute care utilization parameters were clustered into the same segments. Secondly, with additional features representing the types and timing of *post-acute* service utilization, we further partition (with 28-day rehospitalization outcome-supervised machine learning algorithms) the homogeneous segments resulting from clustering into sub-populations characterized by different patterns of *post-acute* service utilization.

Data and features

Included in the current study were records of community-dwelling patients of a municipal public healthcare system in HK (patients admitted from residential care homes were excluded from the current study) who aged 50-64 (n=69,138) and 65+ (n=128,667) admitted between January 1, 2015, and December 31, 2019, and were assigned as their primary diagnoses the ICD-10 diagnostic codes that fell under the different Major Clinical Category (MCC). ICD-10 diagnostic codes, the corresponding procedure codes and the sex and age of the patients, served as the basis from which 218 Case Mix Groups (CMG) of the 20 MCC were constructed.

CMG classifies patients according to clinical principles and can homogenize acute care length of stay for each CMG. However, while patients sharing the same CMG are characterized by their homogeneous diagnostic and procedural profiles, patients' utilization of acute care resources can still vary as a function of their clinical complexity. Clinical complexity is operationally defined by the CIHI Population Grouper Methodology as resource-intensifying comorbidity pairs, cost-intensive interventions, trips to Operating theatres and ICUs, and out-of-hospital operations. The CIHI Population Grouper Methodology estimates acute care utilization for individual CMGs with complexity measures in regression models. Here, the comprehensive list of clinical and service utilization parameters the Canadian public health system applies to derived case-mixed groups from its inpatient population will serve as our population-segmenting parameters.

Analytic models: Unsupervised and supervised machine learning

The community-dwelling patient populations aged 50-64 and 65+ were separately clustered into clinically homogeneous segments to maximize intra-segment similarity and inter-segment dissimilarity according to the following factors with an unsupervised (bottom-up) ML algorithm: Clinical diagnoses, acuity, severity, chronicity, and resource-intensifying comorbidity. We used the unsupervised ML algorithm CLARA (Clustering Large

Applications) an extension of the partitioning around the medoid (PAM) algorithm to achieve the population segmentation. An objective metric was applied to select the optimal number of segments that maximize the within-segment homogeneity and between-segment heterogeneity of the studied population (The Calinski-Harabasz Index). Specifically, CLARA was first applied to the original datasets to randomly create multiple subsets. Second, based on the optimal *k* (i.e., the optimal number of segments) identified according to the Calinski-Harabasz Index, the PAM algorithm was deployed to each subset and identified the corresponding *k* "typical patients" (i.e., medoid, the centre of one segment, the most representative exemplar of patients in the segment with respect to all the clinical and utilization parameters extracted from the populations). Third, using the multidimensional clinical and utilization-related features from CIHI, the PAM algorithm calculated a metric of dissimilarity between each observation of the entire dataset and the medoids and then grouped each observation with the medoid that it was most similar to in terms of the CIHI parameters together as a segment.

Furthermore, to identify patients with the greatest unmet needs compared to other patients with similar clinical status and acute care utilization patterns, a 28-day rehospitalization outcome-supervised machine learning algorithm was applied to further partition each homogeneous segment containing patients with similar clinical and acute service utilization profiles into sub-populations that are also homogeneous with respect to post-acute services that are of high-importance to 28-day rehospitalization. Machine learning models was used instead of the traditional linear models to better describe the high-dimensional and intricately interrelated clinical and service-utilization features, and their unique and combined effects on 28-day re-hospitalization outcomes.

In particular, a non-regression model-based optimal classification tree with an Unbiased Recursive Partitioning and Surrogate Splitting (URPSS) mechanism was implemented. With 28-day rehospitalization as the supervisory outcome, URPSS selected features sequentially in an order that reflected the magnitude of their respective marginal contributions to rehospitalization outcomes regarding every remaining feature in the pool (see Table 1 for the feature pool to which URPSS applied). The greater the magnitude of the selected service's marginal contribution to the 28-day rehospitalization outcome, the earlier it was selected by the URPSS to split the sample of records. Each selected feature (a parental node) splits the remaining records in the segment into two subsets (offspring nodes). On the other hand, the split-stopping criteria of URPSS optimize the tree depth by stopping when the conditional inference of any remaining features in the pool makes no additional marginal contribution to the outcome. One factor of URPSS is that the two offspring nodes that split from the same parental node are statistically significantly different in their outcomes. Hence, inevitably, one of the last sub-populations remained unsplit from the sequential partitioning (i.e., the offspring nodes without turning into parental nodes) would contain only patients who are homogeneous in clinical and acute care utilization AND have received none of the post-acute services that the algorithm selected for its high importance to the rehospitalization outcome. The hypothesis is that, for each segment, the sub-population that received none of the selected post-acute services is always the sub-population whose 28-day re-hospitalization rate is the highest for that segment, even after having adjusted for the conditional inference of all clinical and acute care utilization features. With each no-service sub-population, PAM will again be applied to identify its medoid – the typical patient profile of those who received none of the URPSS-selected services that other patients with similar clinical status and service utilization profiles had nonetheless received.

In the current study, comorbid diagnoses patients received during index hospitalizations were coded separately from the historical chronic illnesses the patients received before the index hospitalizations. In particular, index hospitalization's comorbidity was parameterized according to CIHI's resource-intensifying weights. At the same time, the historical diagnoses of chronic illnesses patients received before index hospitalization were parameterized as presence or absence of six individual chronic illnesses associated with the greatest risk of short-term rehospitalization.

3B. Methods of Studies Motivated by Objective 2

The following is the method adopted by the study reported in the article titled " **Scoring system for predicting the risk of liver cancer among diabetes patients: a random survival forest-guided approach**," which has been submitted to *Liver International* and is currently under revision after review (Yau et al, 2024a).

Study design and study population
A retrospective cohort study of 384,443 patients was conducted using territory-wide electronic health records of Hong Kong's public healthcare system. Under Hong Kong's dual-track healthcare system, the public sector provides approximately 90% inpatient care and 30% outpatient services (including specialist and primary care levels) to the general public. The Hospital Authority (HA) is a statutory body responsible for managing 43 hospitals, 49 specialist outpatient clinics and 74 general outpatient clinics (GOPCs). Since late 2009, the HA has launched a Risk Assessment Management Program (RAMP) in addition to usual diabetes care at GOPCs. Patients who are invited and agree to join the RAMP will receive a diabetes mellitus complication screening (DMCS) assessment to evaluate their cardiovascular profile and risk of developing diabetes-related complications.

The HA maintains a centralized clinical data depository to store patients' demographics, inpatient admissions and outpatient attendances, prescription records, disease diagnoses, and laboratory test results. Disease diagnoses were coded according to the International Classification of Disease 9th or 10th revision (ICD-9 or ICD-10), or the International Classification of Primary Care 2nd edition (ICPC-2). Data were accessed via Hospital Authority Data Collaboration Lab. Ethics approval for secondary data analysis was obtained from the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee.

Patients
Patients who were i) diagnosed with type 2 diabetes mellitus at the age of 18 years or above and ii) referred for a DMCS assessment by a physician at any of the GOPCs during 2010 and 2019 were initially included. Those who had i) a missing date of diabetes diagnosis; ii) a history of malignancy; or iii) less than six months of follow-up were excluded. Patients were followed up until a cancer diagnosis, death, or December 31, 2019.

Input variables
Input variables were ascertained when patients received a first DMCS assessment. Candidate variables for scoring were metabolic indicators (body mass index [BMI], waist-to-hip ratio, glycated hemoglobin [HbA_{1c}], fasting glucose, low-density lipoprotein cholesterol, high-density lipoprotein [HDL] cholesterol, triglycerides, and hypertension), demographics (age, sex, and duration of diabetes), lifestyle behaviors (alcohol and tobacco use), disease diagnosis (chronic hepatitis B/C infection, liver cirrhosis, ischemic heart disease, cerebrovascular disease, heart failure, chronic obstructive pulmonary disease [COPD], pneumonia, chronic kidney disease, and family history of diabetes), medication use (anti-diabetic drugs, aspirin, non-steroidal anti-inflammatory drugs, statins, anti-coagulants, anti-platelets, and anti-hypertensive drugs), biochemical measurements (serum creatinine). Anti-diabetic drugs included were metformin, sulfonylurea, and insulin. In addition, metabolic abnormalities were examined according to international guidelines.

Outcome variable
The outcome of interest was diagnosis of liver cancer (ICD-9: 155; ICD-10: C22) during follow-up.

Data analysis
Due to class imbalance, patients who developed liver cancer (n=2,034) and a random subset of patients who remained free of cancer (n=40,680) during follow up were selected in a 1:20 ratio for model building. Selected patients were randomly split into training, validation, and test sets in a 7:1:2 ratio. Random survival forest was

first applied to rank variable importance and select variables to be included in the scoring systems, and Cox proportional hazards regression was subsequently applied for weight assignment in the scoring systems. A scoring system was first developed among patients of both sexes, and then two sex-specific scoring systems were developed among females and males separately. Variables were selected by considering model performance improvement and model parsimony. Each variable was sequentially added to the initial set of variables in the scoring model according to their ranking in variable importance until no further model performance (area under the curve [AUC]) on validation set was shown. Variables contributing to a relatively large marginal increase in model performance were also additionally considered. For continuous variables, categories for weight assignment were determined based on existing literature, default quantiles, or empirical search. The final scoring systems were fine-tuned based on clinical expert knowledge. The number of trees in the random survival forest was set as 30. Model performance was evaluated using Harrell's concordance (C-) index and AUC as metrics. Comparison of liver cancer-free survival during follow-up across different risk score intervals was conducted using Kaplan-Meier method. Subgroup analysis was performed among patients aged 55 years or above in the absence of chronic hepatitis B/C or liver cirrhosis, where the age cutoff was determined empirically since patients aged 55 years or above demonstrated a markedly elevated risk of liver cancer in the scoring models.

The following is the method adopted by the study reported in the article titled "**Differential metabolic dysfunction profiles and site-specific risk of obesity-related cancers in patients with diabetes**," which has been published by *BMJ Open* (Yau et al, 2024b).

Data source
This is a retrospective cohort study based on territory-wide electronic health records accessed through the Hong Kong's Hospital Authority Data Collaboration Lab (HADCL). The Hospital Authority (HA) is a statutory body providing 90% of inpatient care and 30% of outpatient services to the general population of Hong Kong with more than 7 million people. The HADCL is a centralized repository of public hospitals' and clinics' patient data on demographics, inpatient and outpatient attendances and utilization, disease diagnoses (coded according to the International Classification of Disease 9th or 10th revision [ICD-9 or ICD-10], or the International Classification of Primary Care 2nd edition [ICPC-2]), prescription records, and laboratory results. Ethics approval for secondary data analysis was obtained from the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee.

Subjects
Patients who received diagnosis of diabetes mellitus (DM) in any one of the 43 public hospitals and 73 GOPCs managed by the HA were referred for DM complication screening (DMCS) since early 1990s. All patients who received DMCS were included in the current study with the exception of those who met the following exclusion criteria: i) non-type 2 diabetes; ii) diabetes diagnosis below the age of 18 years; iii) missing date of diabetes diagnosis; iv) history of malignancy; or v) a first assessment before 2010 or after 2015 (as over 95% of patients received a first assessment since 2010, and all patients were followed up until December 2019). In addition, patients who fulfilled the following criteria were also excluded: i) primary cancer diagnosis did not belong to any of the six sites of interest; ii) primary cancer diagnosis at more than one target sites; iii) cancer diagnosis dated within six months or after five years of DMCS; or iv) free of cancer diagnosis but did not survive for more than five years.

Measures
Independent variables were ascertained from the DMCS, which included direct metabolic dysfunction indicators in four categories: 1) obesity (weight, waist circumference, hip circumference); 2) insulin resistance or impaired glucose tolerance (glycated hemoglobin [HbA_{1c}], fasting glucose); 3) serum lipid profile (total cholesterol, LDL, HDL, triglycerides); and 4) hypertension (blood pressure). Also included into the pool of independent variables are those that are not direct indicators of metabolic dysfunction. Nevertheless, these factors may put someone at risk of metabolic dysfunction, such as health-risk behaviours including alcohol consumption and smoking, or

potentially a result of metabolic dysfunction such as disease history (cirrhosis, ischemic heart disease, cerebrovascular disease, heart failure, chronic obstructive pulmonary disease, pneumonia, and chronic kidney disease), medication use (anti-diabetic drugs, aspirin, non-steroidal anti-inflammatory drugs, statins, and anti-coagulants), and laboratory measures (serum potassium, and creatinine; both of which indicate renal function). Patient demographics (sex, birth year, and death year) and family history of diabetes were also included. For individuals who received more than one assessment during the study period, the earliest assessment was considered.

The outcomes of interest were diagnosis of cancer at the colon and rectum (ICD-9:153,154), liver (ICD-9:155), pancreas (ICD-9:157), bladder (ICD-9:188), kidney (ICD-9:189), and stomach (ICD-9:151), within five years after patients have received DMCS. All studied sites are known to be associated with diabetes or obesity. Patients who did not receive any cancer diagnosis served as the reference group in the following analyses. In other words, the studied sample contains seven mutually exclusive groups: six site-specific cancer groups and one reference group who lacked any cancer diagnosis.

Data analysis

Patient characteristics by site-specific cancer outcomes were presented in mean with standard deviation or median with interquartile range for continuous variables, and in count with proportion for categorical variables. Binary logistic regression model was applied to examine the relationships between a set of covariates and each site-specific cancer outcome, among patients who developed a site-specific cancer and those who were free of cancer in any form. Each model included all covariates to mutually control the effects of each other. Estimates of covariates in the models were reported in adjusted odds ratio with 95% confidence interval. Model performance was measured using area under the receiver operating characteristic (AUROC) as metric, with 10-fold cross-validation. All analyses were performed using R (version 3.5.2). Missing values were handled using replacement by attribute mean of the same class. Statistical significance was set at $p<0.05$, two-sided.

The following is the method adopted by the study reported in the article titled "**Comorbidities and prior medical service utilizations affect the mortality trajectories of Chronic Obstructive Pulmonary Disease (COPD) patients: Stratifying 2- and 5- year survival probability of 113,754 community-dwelling 65+ COPD patients discharged from Hong Kong public hospitals with a random survival forest-based decision tool**," and will be submitted to Journal of Clinical Medicine (Leung et al, 2024b).

Sample

The current study examined the 5- and 2-year survival (censored at Dec 2019) of all elderly (65+) COPD patients (ICDs: J410, J411, J418, J42, J430, J431, J432, J438, J439, J440, J441, J448, J449, J684) admitted to Hong Kong's public hospitals (which address the cities' 90% of acute care needs) between 1/1/2007 and 12/31/2012 (n=54,139; hereafter, the 5-year-censor cohort) and between 1/1/2013 and 12/31/2017 (n=59,615; hereafter, the 2-year-censor cohort), respectively. To further ensure the representativeness of the result, a combined cohort (2007-2017) of all above-mentioned 113,754 COPD patients' 2-year survival had also been modeled and compared. Patients' mortality outcomes and corresponding predictors were extracted from patients' electronic health records (EHRs), accessed via Hong Kong's Hospital Authority Data Collaboration Laboratory (HADCL), the centralized data repository of every Hong Kong (HK) public hospital's attendance, covering 90% of the city's overall medical attendance. The project was approved by the Survey and Behavioral Research Ethics Committee (SBREC) of the Chinese University of Hong Kong.

Measures

Mortality outcomes of the studied cohorts were extracted from the territory-wide death registry made available by the Immigration Department. HADCL performed the third-party linkage to ensure the accuracy and anonymity in linking COPD patients' medical records with the same individuals' time and causes of death. Because the death

registry records all known deaths in Hong Kong up until December 31st 2019, this will therefore be our censor time and the basis for sampling for 5- and 2-year-censor cohorts according to patients' admission date.

The primary predictors for mortality outcomes we hypothesized to be of the greatest marginal statistical importance were patients' comorbidities and utilizations at index and previous hospitalizations. Here, in terms of comorbidities, we included into the feature pool 1) the number of times the utilization of acute care resources be intensified due to specific comorbiding diagnoses the patients received at index hospitalization according to CIHI, 2) having histories of any of the five frailty-related chronic conditions (according to the Mayo clinic) as comorbidity, and 3) all the secondary diagnoses patients had received prior to, and at, index hospitalizations. On the other hand, metrics of medical service utilizations were included: past-year cumulative counts of hospitalization and cumulative length of stay, as well as the index hospitalization's length of stay.

In addition to comorbidities and utilization, factors such as patient's age and sex, and biomarkers representing clinical acuity and severity were also included in the predictive feature pools such that their respective effects on survival outcomes could be adjusted in our analysis of comorbidities and utilizations' importance to survival outcomes. Biomarkers for COPD patients' lung functions and general clinical acuity were both examined. In terms of biomarkers indicative of patients' lung function, such as the patients' GOLD stage, partial pressure of oxygen and carbon dioxide, pH levels in the blood were included as predictive features due to their associations with COPD patients' mortality reported in the literature. In addition, biomarkers that were not directly linked to lung functions but were nonetheless associated with the clinical acuity and mortality outcomes of COPD patients in previous studies were also examined here, such as hemoglobin, white blood cells count, and blood glucose levels, as well as blood urea nitrogen, serum creatinine, and serum potassium. Notably, given that an increased likelihood of adverse outcome is associated with the above-mentioned biomarkers' levels being either above or below the normal range, the biomarkers were engineered into three response levels before being included in the model and each level's marginal effect on survival outcome was quantified.

Analytic Models

The current study applied a random survival forest (RSF)-based algorithm to rank order features according to their respective importance to survival outcome, categorize variables and assign weights to optimize outcome prediction, identify and fine-tune the cutoff derived from the cumulation of assigned weights to optimize outcome prediction, and validate the weight assignment and cutoff identified.

First, we applied an RSF algorithm whose objective was to rank order the predictive features according to their importance to survival outcome. RSF consists of a multitude of survival trees grown by a recursive splitting of tree nodes to maximize survival difference (log-rank test statistic) and estimate the survival probabilities based on the ensemble cumulative hazard function. To enhance the accuracy all for model prediction and variable ranking over individual survival tree, RSF applied an ensembling process to combine the results of different survival trees, bootstrap different samples of data, and integrate different randomly selected subsets of variables.

Second, feature importance is estimated from a corresponding drop in predictive accuracy when the value of one variable is replaced with its random permutation value. As the determinants of individual features' contributions to overall predictive accuracy, the coefficients of RSF are different from that of the traditional Cox regression, whereby the RSF doesn't assume proportional hazard or any functional form for the hazard function. RSF also out-performs traditional Cox regression because it works well for high-dimensional EHR data.

Third, based on the magnitude of their importance, features were selected and selected features preprocess for transformation. For example, continuous variables were transformed into categorical ones according to the value's quantiles (e.g., 0%, 5%, 20%, 80%, 95%, 100%). Consequently, every response level of each of the selected feature was assigned a partial (marginal) weight derived from the resulting coefficients through a two-step procedure whereby the reference category we assigned the smallest coefficient, and then the coefficients assigned to the non-reference response level were divided by the reference coefficients and rounded to the nearest

integer. Consequently, a time-to-mortality score was computed from aggregating all weights assigned to every response level of each selected feature.

To characterize how well the time-to-mortality score distinguished between subjects who died within the censor period and those who survived beyond it, we applied Kaplan-Meier estimator of the survival function and assessed the score's discriminatory performance as a function of time using AUC(t) – which is defined by cumulative sensitivity and dynamic specificity, as this definition has more clinical relevance and has commonly been used by clinical applications. To obtain a single overall performance metric to guide the parsimony in feature selection, we derived the integrated AUC (iAUC), a weighted average of AUC(t) over specific follow-up periods (i.e., from Day 1 to Day 90), summarizing the overall discrimination ability of the time-to mortality score.

Model performance was reported on the validation dataset, and 100 boot-strapped samples were applied to calculate 95% confidence intervals (CI). The model's discriminatory performance was considered poor if AUC<.70, acceptable when AUC=.70 to <.80, excellent when AUC=.80-.90, and outstanding when >.90.

3C. Methods of Studies Motivated by Objective 3

Please refer to https://doi.org/10.3390/ijerph21020179 for the method adopted by the study reported in the published article titled "**The environmental, sociodemographic, and clinical determinants of pain interference among the working poor: The development and validation of a random forest-based stratification tool on the socioecological impact**" (Leung et al, 2024c). The following is the method adopted by the study reported in an forthcoming article on using a supervised learning-abbreviation of the Comprehensive Geriatric Assessment (CGA) titled "**Screening for Frequent Hospitalization Risk among Community-dwelling Elderly between 2016 and 2023: Machine learning-driven item selection, scoring system development, and prospective validation**." to appear in *Frontier in Public Health* (Leung et al, 2024d).

Samples and Data collection
Between 2016-2018, nurses and social workers administered the CGAs in pairs at the homes of 1611 elderly (65% female) aged 65+ (mean=80.51, SD=7.1). This cohort of 1611 elderly were recruited from the clientele of one type of government-subsidized community service whose mandate is to provide those who can live independently in the community (as judged by licensed social workers) with a network of support to provide care and concerns when needed. In addition, to ensure generalizability of our short-form CGA, we have administered it to a prospective validation (28) cohort of 329 elderly (69% female; mean age 76.35 (SD=8.02), recruited between 2022 and 2023 from a series of community health assessment events held at the community centres in a district adjacent to the ones from which we sampled our 1611 participants.

The development and validation of short-form CGA in a cohort of 1611 elderly clients
Operational Definition of Frequent Hospitalizations
In the current study, frequent hospitalization outcomes were derived from participants' self-report of the dates of their past hospitalizations over the two years prior to when CGAs were conducted.

Specifically, as previous studies had shown, the closer the hospitalizations were to when the assessment of care needs was performed, the greater the care needs the individuals had, even compared to those hospitalized with the same frequency but spread over the entire study period. Hence, not only did the current study operationally define frequent hospitalization as two or more hospitalizations over a two-year period in accordance with the literature on frequent hospitalization, but it also examined different operational definitions with respect to whether the self-report hospitalizations occurred in the year immediately before when the participants were asked to recall the hospitalization events (namely "the 2nd year") or did the hospitalization occur during the year prior to the one immediately before when the assessment was conducted ("the first year").

Here, depending on whether hospitalizations occurred during the first or the second year (or both), four operational definitions of frequent hospitalizations were derived from a progressively less inclusive temporal distribution of their two or more hospitalization events. In particular, our first and most inclusive operational definition of frequent hospitalization requires only two or more hospitalizations to occur at any time over the two-year prior to when the assessment was conducted. On the other hand, the second and incrementally more restrictive operational definition required that at least one hospitalization occur in year one. Even more restrictive was the third operational definition, which included the scenarios of 1) having two or more hospitalizations in year one, regardless of year two's hospitalization pattern or 2) having exactly one hospitalization in each of the two years. Finally, the most restrictive operational definition of frequent hospitalization studied here required that two or more hospitalizations occur at year two, regardless of one's hospitalization pattern during the 1st year.

Dimension Reduction of CGA with Machine Learning and Deep Learning Algorithms

The purpose of deriving different definitions from scenarios involving two or more hospitalizations over a two-year period is to compare across the four operational definitions the following: 1) the performance of LASSO, non-regression-based decision tree, AdaBoost, and DeepBoost algorithms and 2) the identity and order of items selected by the best-performing algorithm during learning model-supervised dimension reduction. On the one hand, performance of the studied algorithms is parameterized, and compared, in terms of the area under the receiver operating characteristic curve (AUC). The model's discriminatory performance was considered poor if Area Under Curve <.70, acceptable when AUC=.70 to <.80, excellent when AUC=.80-.90, and outstanding when AUC >.90. On the other hand, Spearman's rank correlations were performed to compare the items selected, and the order in which items were selected, by the best-performing algorithm's four operational definitions of frequent hospitalization outcomes were examined.

Risk scoring the short-form CGA with a random forest-based algorithm

For the CGA items selected by the best-performing learning algorithm to achieve their "intended clinical use" in screening community-dwelling elderly for frequent hospitalization risk, outcome differentiability-optimizing weights were assigned to every response level of each selected item by a random forest-based algorithm to create an interpretable and implementable scoring system. Specifically, the weights constituting the scoring system are developed from the coefficients that the random forest component of our algorithm assigned to every response level of each feature in accordance with each response level's contribution to the frequent hospitalization outcome. Subsequently, the response levels assigned the lowest value of coefficients were designated as the reference category against which coefficients of other non-reference response levels were divided and rounded to the nearest integer to derive their corresponding score values. Finally, the scores that correspond to each response level, and the cutoff value beyond which the sum of response-level scores is said to represent frequent hospitalization risk, were derived to maximize outcome-differentiability. In the current study, only the items selected by the best-performing algorithm will be developed into a scoring system. In addition, only the most inclusive operational definition of frequent hospitalizations was used in the construction of the scoring system (i.e. the first operational definition: Two hospitalization events over the course of two years), due to the low expected frequent hospitalization risk among community-dwelling elderly in the general population.

Prospective validation of the scoring system developed from the short-form CGA with a cohort of 329 elderly during the COVID-19 pandemic

To ensure the generalizability of our short-form CGA and the scoring system from which it derived, we performed a prospective validation under conditions different from those under which the CGA was initially abbreviated and its scoring system devised. Firstly, the validation sample was recruited from a district adjacent to the original population where the clientele of the government-subsidized community service was sampled to develop the short-form CGA. Secondly, members of the general population who attended health assessment events hosted by the participating NGO, rather than clients of the government-subsidized community service, were recruited via venue-based sampling. Thirdly, the scoring of the validation cohort's frequent hospitalization risk was performed by community workers and healthcare professional trainees, commonly found in NGOs of the local context,

instead of licensed nurses and social workers who initially performed the CGA in the home of the 1611 clients of the government-subsidized community service. Finally, the validation cohort was recruited during the COVID-19 pandemic between 2022 and 2023, rather than during the same period when the initial cohort of 1611 was recruited between 2016 and 2018 to develop the short-form of CGA.

In addition, members of the validation cohort are considered at risk of frequent hospitalization if they were scored above the cutoff value of the scoring system. On the other hand, their self-report frequency of hospitalizations during the two years prior to the date of risk scoring was ascertained as the outcome for validating the at-risk status scored by our risk-scoring system. A logistic regression model was applied to validate the at-risk status scored against the self-report outcome. The performance of the validation model was measured using the AUC.

Please refer to https://doi.org/10.1038/s41598-022-25924-6 for the service optimization algorithms developed and the method adopted by the study reported in the published article titled " **Robust meal delivery service for the elderly: a case study in Hong Kong**" (Lin, S., Bakker, M. & Leung, E. (2022). *Scientific Report*)

The following is the service optimization algorithms developed and the method adopted by the study reported in a forthcoming article titled " **Data-driven optimization of a population-level allocation of post-discharge services for diverse patient segments to improve patient outcome and maximize medical cost savings: A case for decision-centric precision public health**," to appear in *Frontier in Public Health* (Leung et al, 2024e).

The current study is based on secondary analyses of a previously published cohort of community-dwelling patients aged 65+ (n=128,667) admitted to a municipal public healthcare system in HK between January 1, 2015, and December 31, 2019. The studied cohort was segmented into clinically homogeneous patient segments by an unsupervised learning algorithm and, within each patient segment, further partitioned by applying 28-day rehospitalization-supervised machine learning algorithms to acute and postacute service utilization features to identify the subpopulations of patients who lacked the postacute services that were nonetheless received by other patients sharing homogeneous clinical and acute service utilization profiles (hereafter, the "no service" group, or NS in short; for detailed descriptions of the cohort and the unsupervised and supervised-learning models applied, please refer to Leung et al 2004a). Here, the NS group served as the counterfactual scenario (reference group) in our generalized cost-effectiveness analysis (GCEA) with which each patient segment's post-acute services are paired to form a "GCEA pair," and based on which a linear program was performed.

Specifically, we performed the following calculations for the purpose of GCEA: 1) extract the cost data on acute (bed days), ambulatory and postacute services that that HK public medical system has made public, and calculate the care episode's cost of all ambulatory, acute (bed days), post-acute care, including cost associated with re-hospitalization within 28 days of discharge; 2) compare among patients who shared homogeneous clinical and acute service utilization profiles the sum of care episode cost between each GCEA pair, i.e. between those who received one type of ambulatory and postacute service selected by our supervised learning model and those who shared the same profile but did not receive any selected ambulatory and post-acute services (the NS "counterfactual scenarios") – the difference was considered in the current study as the net cost savings of the type of ambulatory or postacute service studied in each pair; and 3) parameterize a GCEA ratio to reflect the additional amount of health system-wide (across acute, ambulatory, and postacute care) cost saving for every additional dollar spent on each selected ambulatory and postacute care received by each patient segment. The GCEA ratio is calculated from the regression coefficients of the regression lines extrapolated from system-wide cost saving every year between 2015-2019 and spending on ambulatory and post-acute services every year between 2015-2019. Negative and statistically significant regression coefficients (i.e., the slope of the regression line being negative) suggest that system-wide cost saving is associated with an additional unit (dollar) of spending in ambulatory and post-acute services.

Finally, with the regression coefficient of the relationship between system-wide cost saving and individual service spending as the GCEA ratio, a linear programming model was developed from it to identify the system-wide cost-optimal pattern of allocating ambulatory and postacute services in each patient segment of homogeneous clinical and acute care utilization. Please refer to the Appendix below for the formation and a detailed description of the linear programming models. The objective function of the linear programming model is to maximize the reduction of acute, postacute, and ambulatory care costs associated with 28-day rehospitalization. Based on this objective function, two scenarios were programmed from two different sets of constraints. Under Scenario One, the constraint was that the total spending on medical costs, including those associated with 28-day re-hospitalization, must not exceed the current level. Even though under Scenario One, the spending on individual ambulatory and postacute services can be increased by the same amount that could potentially be saved in terms of reduction in 28-day re-hospitalization if the NS groups had received the selected ambulatory or postacute services. On the other hand, the constraint of Scenario Two was that neither the total cost nor the spending on specific ambulatory or postacute services was allowed to increase. Only the paring between patient profiles and the types of ambulatory or postacute services is allowed to change. The more stringent constraints in Scenario Two is meant to reflect that the limit of service capacity may not always be increased in spending even if it came from cost savings rather than additional injection of fund, such as a shortage of healthcare professional.

The following is the method adopted by the study reported in the article titled "**The Valuation of Elderly Homecare Services Under a Joint Medical-Social Budget Perspective**" which is forthcoming in *Frontier in Public Health* (Leung et al, 2024f).

Study setting

The current study consisted of 633 clients of a homecare service who consented to the service receiving their discharge summaries issued by hospitals in the same catchment area. The studied homecare was a standard service delivered by one of the sixty NGOs commissioned by the social care bureaucracy of Hong Kong (HK) 's publicly funded healthcare system. It is of note that HK's healthcare system also provides medical care under a separate bureaucracy. The mandated objective of the studied homecare services is to facilitate "aging in place" among clients who are otherwise capable of living independently in the community despite transient unfavorable circumstances. At intake, a licensed social worker assessed the clients' needs, based on which the clients were assigned one or more types of services of a personal, instrumental, and restorative nature.

Personal homecare offered assistance with bathing and general domestic duties. Instrumental homecare assisted with shopping (e.g., purchase and delivery of daily necessities), food preparation (e.g., provision of meals), transportation (e.g., escort to clinics), managing finances (e.g., application of financial assistance), and housekeeping. Finally, restorative homecare managed and prevented different clinical issues, including accidents and falls, wounds and pressure injury, medications, diet and nutrition, cognitive impairment, convalescent, infection control, chronic pain, depression, agitation/aggressive behaviors, constipation, incontinence, etc.

All participants were assigned at least one homecare service. However, not everyone's service package had commenced within the study period. As a result, some participants have no recorded service transactions on file.

Data and variables

Data extracted from the hospital discharge summaries included the timestamps, types, and duration of each medical encounter and the client's corresponding medical diagnosis. On the other hand, data extracted from homecare clients' records were timestamps of when a homecare service was received, and the type(s) of homecare services received at each transaction. Each client's data resulting from their medical and social care were linked and sequenced together with respect to the chronology of timestamps. Consequently, a dataset was created to host data generated at both the inter-individual and intra-individual levels.

The outcome variable was the time between consecutive medical encounters (for example, days between the convalescent care from which a client was discharged and the subsequent acute care to which a client was admitted). Post-discharge records were right-censored if no subsequent admission to medical care was found. Explanatory variables were the types of homecare services the client received between consecutive medical encounters (the service's duration and frequencies were used for calculating the total homecare cost, see below). To adjust confounds from estimating each type of homecare service's effect on the number of days between consecutive medical encounters, we included the following covariates: The clients' demographics, medical history of chronic illnesses, licensed social worker-assessed functional impairment, emotional and cognitive issues, and the presence of chronic risk factors such as smoking. Also adjusted at covariates were the types and intensity of the acute and post-acute follow-up services utilized.

Calculation of medical and homecare costs

The cost of each medical encounter of a client was calculated by totaling the published cost of each service utilized during this medical encounter. For example, the published average cost of each A&E attendance was HK$1,230 (US$159; exchange rate of US$1 to HK$7.75), each bed-day at the acute inpatient ward was HK$5,490 (US$708), and each bed-day at a convalescent hospital was HK$2,390 (US$308) in 2015/16. Consequently, the medical cost for a homecare client who spent three days at an acute hospital followed by two days at the convalescent hospital is HK$22,480 (US$2,899). To align with homecare costs, we adjusted the medical costs with the inflation rate using the Composite Consumer Price Index of the Census and Statistics Department of HK.

The unit costs of each of the three homecare services have not been made public by the social care bureaucracy. Hence, we estimate the unit cost of each service type from two components: One that is unique to each service and one shared by all three service types. Service-specific costs were calculated from each professional category's published median hourly rate to which service-specific staff belonged and the number of hours spent delivering each service. Shared costs were estimated by deducting service-specific costs from the total homecare budget, derived from the HK government's annual budget allocated to service operators for each homecare client across sixty NGOs.

Analysis

With the number of hospital-free days spent in the community since the last hospital discharge as an outcome, Cox's proportional hazards model was built from the utilization metrics of different types of homecare services received between medical encounters as explanatory variables. In addition, our modeling of hospitalization outcomes from homecare utilization was adjusted for the effects of the following covariates: Clients' medical/functional characteristics and the types of medical care received since last discharge. In addition, a gamma frailty term[28] was added to Cox's model to account for the effects of individual differences in deterioration rate on hospitalization outcome. Consequently, Cox's frailty model estimated, in terms of risk-adjusted hazard ratios (HR), each homecare service's marginal "survival benefit," expressed as the reduction of hospitalization risk by an amount equal to 100*(1-HR)%.

Furthermore, log-likelihood ratio (LR) tests were performed to demonstrate that Cox's frailty model was superior in data fitting to one without the frailty term or a null model with no explanatory variable when modeling the "survival benefit" of elderly homecare services. In addition, to assess the performance of the predictive models and their respective features, concordance statistics (c-statistics) of the Cox's frailty model and the single-level Cox model were computed respectively and compared. A c-statistic between 0.60 and 0.70 indicated fair to modest performance, and a concordance of 0.80 and above indicated excellent performance.

Several properties of Cox's model and the resulting HRs are especially relevant to our valuation of homecare's complex interventions under a GCEA framework. For example, in the counterfactual scenario that serves as the reference through which different combination of homecare services' effects were compared to one another, the subgroup that received no homecare service was assigned an HR of "1" and thus served as the baseline for risk

reduction (i.e., 100*(1-HR)%=0%) to which those who received different compositions of homecare were compared. In addition, the reference value of HR=1 could be assigned to any service combinations of homecare and be chosen the counterfactual scenario, depending on the services targeted for comparison in the GCEA.

Another property of Cox's model that adds value to GCEA is HR's multiplication property. To operationalize the combined risk-reduction effects of the multiple services of homecare's complex intervention in relation to its baseline scenario, *r* is therefore advanced here as the product of between the risk-adjusted HR of a composition and its baseline HR, which may be something other than one if the counterfactual baseline consists of HR >= 1.

Hence, in addition to the no-service (or selected-service) counterfactual scenario, whose HR was the reference value from which the corresponding service compositions' *r*s were calculated, each homecare composition also has its counterfactual scenario in estimating its related medical costs. The counterfactual medical cost refers to the medical cost that recipients of specific homecare composition would have incurred if they had not received the assigned homecare composition. Hence, the counterfactual medical cost (M_a) of each homecare composition was calculated by adjusting off homecare's reduction of hospitalization risk (*r*) from its associated observed medical cost (M_0), i.e., $M_a=M_0/r$. Consequently, medical cost saving is noted ($\Delta M>0$, where $\Delta M=M_a-M_0$) when the risk of hospitalization outcomes is reduced among those who received homecare relative to those who did not ($r<1$). On the other hand, no medical cost-savings ($\Delta M=0$) could be concluded if $r=1$ and additional medical costs had been incurred ($\Delta M<0$) if $r>1$.

Finally, to examine net cost-savings from a joint medical-social budget perspective, we developed a ratio representing homecare's medical cost savings relative to the cost of homecare. Specifically, homecare cost was calculated by first tallying the unit cost of each homecare service offered every time during medical encounters (t), then taking the median of all between-medical-encounter total costs. Since not all counterfactual medical costs were calculated with reference to a scenario that lacks any homecare service, the net cost-saving ratio presented here is calculated by dividing from the medical cost savings (i.e., ΔM) the difference between the counterfactual scenario's homecare cost and the cost incurred from specific homecare composition of interest ($\Delta S(t)$). Consequently, the ratio of net cost saving is $E=\Delta M/\Delta S(t)$.

In addition, notwithstanding the clients were also recipients of other homecare services concurrently, restorative homecare was generally assigned to those suffering from higher medical and social needs. Hence, GCEA was performed separately for those who received restorative homecare and those who did not.

3D. Methods of Studies Motivated by Objective 4

Three studies were conducted to achieve the fourth objective of establishing a novel methodological framework and a technological platform to enable data-driven primary care policy design and facilitate the expansion of Hong Kong Smart City solution of Smart Environment to primary care. The resulting novel methodological framework and a technological platform developed from the three studies were adopted to support the work of an NGO, a Municipal Government, and the Hospital Authority, respectively.

Hong Kong is striving to become a smart city, bringing innovative and transformative technologies to its healthcare system. However, its public primary care system has been underinvested, and its public housing, which accommodates almost half of Hong Kong's population, has outdated design elements that were known health hazards and contributed to the SARS outbreak 20 years ago. With research linking both the internal and external built environment to COVID-19 outbreaks at the pandemic's onset, and with us a peri-COVID era where occasional yet persistent outbreaks are the new normal, anticipating when and where COVID-19 outbreaks will occur and their impact on our medical system has become a primary care priority.

Hence, three studies were conducted to improve our ability to forecast COVID-19 outbreaks and their impact on the medical system. The aims of the three studies were achieved using a novel deep learning methodology designed for multi-source data analysis to align with the multi-layered socioecology of each studied district.

The first study investigated whether the internal built environment had affected residents' general health before SARS in 2003, and if the same elements in the built environment were also responsible for increasing the residents' susceptibility to COVID-19 during the pandemic's first year. We have also examined if our prediction model can be replicated in other districts with similar public housing designs.

The second study broadened the scope to examine the buildings' sociodemographic profiles and their internal and external built environment. The aims of the second study were to identify and compare elements of greatest statistical importance in forecasting COVID-19 outbreaks over 14-day horizons during each pandemic wave across the three studied districts.

The final study adapted our forecasting model for COVID-19 case count to include time-varying case count and surge history, as well as time-invariant features such as the buildings' sociodemographic profiles and their internal and external built environment to forecast surges at emergency departments (i.e., ED boarding for 4+ hours, which has been shown to adversely affect patient outcomes). Our model forecasted ED surges during the pre-COVID-19 period, across all five waves of COVID-19, and within the brief period between waves four and five. This brief period, when the residential population co-existed with the dormant virus, is significant as it resembles our current period of "living with the virus." The third study also examined the transferability of the forecasting model built from each wave or period to all other waves and periods to determine which waves or periods' perturbation to our medical system were most revealing of the underlying patterns of interconnectedness among the elements in our medical system.

While the three studies had different focuses, they shared the same methodological concern over ensuring that our deep learning algorithm is ecologically valid in terms of its alignment with expert knowledge, the multi-source nature of the real-world data involved, and the intended clinical use or policy decision support of the algorithm. Specifically, we have "anthropomorphized" an AI algorithm according to the hierarchical organization of different elements in a person's socioecology previously linked to COVID-19 or poor health more broadly. To ensure ecological validity, we developed a hierarchical input architecture using features sourced from multiple public databases. This AI algorithm, with its hierarchically organized and closely interrelated features, was used to predict the case counts for 360 buildings accumulated during the pandemic or ED surges before and throughout the pandemic. The model's validity was then confirmed by its forecasting performance over rolling 14-day horizons within a two-month period.

For details on the deep learning methods reported in studies motivated by objective 4, please refer to the manuscripts Guan et al (2023), Guan et al (2024), Leung et al (2024g) and Leung et al (2024h).

A more detailed account of our algorithm's architecture is also presented in the following. The development of COVID-19 case counts- or ED surge history-supervised Multi-Headed Hierarchical Convolutional Neural Networks (MHHCNNs) was based on the studied buildings' internal and external built environments and the residents' sociodemographic profiles. Each input layer was connected to an embedding layer followed by a long short-term memory (LSTM) layer. While each input layer was comprised of all levels of a single feature, different layers representing features belonging to the same socioecological levels (building-level, estate-level, and Tertiary Planning Unit Group- (TPU-) level) were concatenated and served as input to one convolutional layer. We expected that the performance of modeling COVID-19 case counts or ED surge history would be enhanced by including elements of buildings' internal and external built environments and the residents' sociodemographic profiles, as all input features were hierarchically structured in alignment with the different levels of the

socioecological framework. Performance is measured in terms of the area located under the receiver operating curves (AUCs) and epoch to the MHHCNN model convergence.

In addition, MHHCNN also assigned a measure of importance (the Shapely value; see below) to each feature according to its unique contribution to COVID-19 case counts or the proportion of the day with ED surge accumulated over the period being modeled. Notably, the unique contribution of each feature is estimated within the context of the contributions of all other features and their interactions. In particular, the hierarchically organized concatenation of the MHHCNN also allows the importance associated with interaction among features from the same or different input layers to be estimated in isolation within the context of other features' contributions, singly or in combination.

Shapely value is a method based on cooperative game theory and has been applied to increase the transparency and interpretability of machine learning models. Shapely value is a numerical expression of the marginal contribution of each parameter to the outcome (i.e. feature importance). The unique contribution of each parameter can be expressed as the degree of change in overall performance when the parameter is excluded. Shapely value is more sensitive, consistent, and accurate than the standardized regression coefficient, which is also aimed at parameterizing unique contributions of individual features but does so in linear models. Notably, SHAP has been dubbed as the key tool in the explainable artificial intelligence's approach to making deep learning models more interpretable to end-users. In the current study, a Shapely value was assigned (by MHHCNNs) to each feature value of each feature modeled as a predictor.

It is a well-established practice for statistical- or machine-learning models to be cross-validated by dividing the entire dataset into k subsamples, with k-1 subsamples being the model-building set and the remaining one as the model validation set. The random subsampling of model-building and model-validation sets and the subsequent validation of models will be repeated k times. No dependency nor systematic difference is expected between the model-building and model-validation sets. However, dependency and systematic differences between model-building and model-validation sets can be expected if one were to sample the model validation sets in subsequent periods to align with the "intended clinical use" of risk prediction models, which is to predict future risk. Predicting COVID-19 case counts accumulated or ED surge over time is challenging. On the one hand, the growth of COVID-19 case counts and the proportion of day with ED surges is non-linear and affected by factors that apply to the entire population, such as the virology of the virus, as well as more heterogeneous factors, such as different individuals' socioecology. On the other hand, the growth of COVID-19 or the surges in ED are also affected by the different public health and social measures put in place with an intention to curb the spread of the virus at various times in response to the changing case counts. Hence, we have validated our model with data extracted during the two-month period subsequent to when model-building data were extracted with the forward-chaining validation methodology.
The data extraction period of the first chain's training phase spanned between 2021-12-24 and 2022-05-21, while the testing phase's data extraction period was between 2022-05-22 and 2022-06-04. The training phase of the second chain was based on data extracted over the period that overlapped with both the training and testing phases of the first chain, and the second chain's testing phase was based on data extracted from the 14 days that followed (from 2022-06-05 to 2202-06-18). By the same token, the third chain's training period overlaps with the period when the previous chain's training and testing phases had taken place, with its own 14-day testing phase ending on 2022-07-23. During each testing phase, the performance of forecasting over the 3-, 7-, and 14-day horizons was compared between LSTM deep neural network (DNN) models consisting of both the composite Shapley value of building features and the historical COVID-19 case counts or ED surges recorded during the training phase and LSTM DNN models that consisted of only historical COVID-19 case counts or ED surges over the same period.

In addition to prediction modelling and forecasting, deep learning model was also deployed here for testing transferability of models (i.e. the underlying interconnectedness of different healthcare element represented by the period when the model was built).

Machine learning algorithms are traditionally designed to train every model in isolation based on the specific domain, data, and task to solve specific tasks. The models have to be rebuilt from scratch once the feature-space distribution changes. Transfer learning is a means to overcome the isolated learning paradigm and generalize the knowledge acquired for one task to solve related one. For this reason, the extent to which a model could be transferred across domains and data reflects the relatedness of the domains and/or data in question. And it is the transferability of the learning model built between waves four and five that the current study sought to test.

The machinic of testing the transferability of models with transfer learning is straightforward. In machine learning, a large amount of training data is typically required for training a model from scratch. Transfer learning generalizes the machine learning model developed from one context (domain) to another despite large amounts of training data or data points that are common across the two domains which are not available. Specifically, transfer learning enables us to utilize knowledge from previously learned tasks and apply them to newer, related ones. If we have significantly more data for task T1, we may utilize its learnings and generalize them for task T2 despite T2 having significantly fewer data. Compared to the ignorant learner model, transfer learning can improve baseline performance, reduce the overall amount of time taken to develop/learn, and improve the final performance of a model. Transfer learning creates an inductive bias on how and what is learned by the new model in performing a similar task within the context of a different domain. It does so by narrowing the hypothesis space from which the new model is developed, and facilitating the process with which learning is performed within the hypothesis space.

The machine learning model has highly configurable architectures that are layered with parameters. These layers have initial layers designed to capture the more generic features and later ones that focus on the more specific task at hand. These layers are normally connected to each other and to the final classification layer. The final layer is in turn trained by specific output. This layered architecture allows us to remove the final classification layer and create a “pre-trained” model to select features with respect to a similar task in a new domain. Transfer learning is the mechanism for creating pre-trained models from existing machine learning models in order to generalize the knowledge (in terms of model parameters and weights extracted of the existing model) to inform feature selections of the new model for a new task in a new domain. As a starting point of developing a new model in a new domain, transfer learning extracts the layered weights and parameters from an existing machine learning model and fine-tune them to enhance the model- training process in a new domain by improving the baseline performance, efficiency gain and final model performance.

CHAPTER 4 RESULTS

In this section, an overview of the scientific findings emerging from the current project, elucidated in accordance with the logic of the multi-source data analysis performed under the CDC's socio-ecological framework for prevention. The scientific findings were reported in full in manuscripts enclosed in the Appendices previously referred, and are currently either being submitted, forthcoming, or already published. Being the academic output of the current study to date, the manuscripts enclosed in the appendices are the basis of the recommendations provided below and the journey of knowledge translation/dissemination we embarked on with our stakeholders.

4A. Findings reflecting Objective 1

We began the project with studies conducted to achieve Objective #1 at the person-level of the residents' socioecology using K&T and SSP outpatient and inpatient population EHRs to identify the "typical patient profiles" who showed high 28-day rehospitalization rates. The two districts were the same for the "typical" high-needs patient profiles that were stable over a five-year period: Patients with COPD (especially those who aged between 50-64, who suffered from higher acuity, chronicity, complexity, and showing a shorter time between discharge and rehospitalization), and patients who suffered from the cancers of the digestive (particularly for those who aged 65+), nephrology and urology (especially for those who aged 50-64) systems. Unique for residents of K&T were typical profiles of pain-related hospitalisation and mental health-related patient profiles.

A paper "**Article- Driving precision public health research and intervention in Hong Kong with population data: A machine learning-informed joint medical-social perspectives" has been submitted to Hong Kong Medical Journal** under review (**Leung et al, 2024a**).

In that study, the following seven systems of diseases and disorders were identified as the major clinical categories of the medoids around which the 65+ and 50-64 populations were partitioned: Circulatory, Digestive, Nephrology and Urology, Musculoskeletal, Respiratory, Multi-system, and Other Reasons for Hospitalization. Mental Diseases and Disorders was identified for the 65+ population as the eighth segment. The different clinically homogeneous segments resulting from the partitioning around their respective medoids will be named according to the medoids' major clinical categories by mapping of the International Classification of Diseases (ICDs), Case Mixed Groups (CMGs) and Major Clinical Categories (MCCs) constituting each segment and its corresponding medoid.

Each segment's sub-population receiving none of the Unbiased Recursive Partitioning and Surrogate Splitting (URPSS)-selected post-acute services (hereafter referred to as no service group, "NS") consistently showed a higher 28-day re-hospitalization rate and a shorter median time between discharge and re-hospitalization (most medians were less than 14 days) compared to the corresponding population segment at a whole. The NS groups of the 65+ and 50-64 populations differed in their patterns of resource-intensifying comorbidity and interventions (i.e., cardioversion, chemotherapy, dialysis, cell saver, heart resuscitation, paracentesis, pleurocentesis, radiotherapy, non-invasive biopsy or per orifice endoscopy) observed at index hospitalization, and the percentage of being diagnosed with two or more chronic illnesses. The NS groups of all 50-64 segments showed a higher percentage of being diagnosed with comorbidity that intensified acute care resource utilization by two or more times and having resource-intensive intervention during index hospitalization, and a higher percentage of being diagnosed with two or more chronic illnesses, compared to their corresponding full segments. While the 65+ population's NS groups also consistently showed a higher percentage of being diagnosed with two or more chronic illnesses compared to their corresponding full segments, any discernible patterns were not observed in terms of the prevalence of resource-intensifying comorbidity or flagged interventions.

The 28-day re-hospitalisation rates of the 50-64 NS groups were found to be greater than that of the 65+ population, with the exception of the circulatory and musculoskeletal segments. Higher percentage having

comorbidity incurring two or more times the acute care resource utilization and resource-intensive flagged interventions were found among the 50-64 NS groups who exhibited greater 28-day re-hospitalization rates than their 65+ counterpart.

Among those who received URPSS-selected services in both populations, ambulatory follow-up care at specialist outpatient clinics (hereafter referred to as SOPC follow-up) was consistently the first to be selected across all segments. Being first selected in the URPSS process reflected that SOPC follow-up's unique contribution to 28-day re-hospitalisation outcomes was the greatest among all the selected services, whose contributions to outcomes were conditional upon SOPC follow-ups being absent. Patients who received SOPC follow-up consistently exhibited the lowest 28-day re-hospitalization rates compared to their clinical and acute care-need homogeneous counterparts in their respective segments who received other services post-discharge, with the exception of 65+ and 50-64's Digestive and Respiratory segments and the 50-64's segment of Other Reasons for Hospitalization, the second lowest 28-day re-hospitalization rates were associated with SOPC follow up and the lowest ones were associated with receiving nursing transitional care delivered to the home of community-dwelling patients or visiting primary care clinics.

Even though SOPC follow-up was associated with the lowest 28-day re-hospitalization rates in most cases in both 65+ and 50-64 populations, the average wait time for SOPC appointment ranges between 9 to 111 weeks (https://www.ha.org.hk/haho/ho/sopc/dw_wait_ls.pdf), and the median time between discharge and rehospitalization of NS patients is 14 days. If SOPC services were not available to those in the study population, other services would still help to reduce re-admission but the extent differed for different segments and not substantially for most of the services operating in silos. In this study, the partition of the population is optimised to reveal the population's unmet care needs and the corresponding service gaps. The in-patient population aged 50 and above, 28-day re-hospitalisation rate was over 80% when no machine-selected services were received, i.e., the NS groups. On the other hand, when compared among those who shared similar clinical and acute care utilization profiles, those who received SOPC were associated with significantly lower rehospitalization rates compared to those who received any other machine-selected service type instead.

The extent to which the waiting time for SOPC appointments to be shrunk is constrained by the dwindling availability of healthcare professionals. The alternative services our medical system made available to post-discharge patients were consistently associated with a higher re-hospitalisation rate compared to patients with similar clinical and service utilization profiles who received SOPC, but their re-hospitalization rates were significantly **lower** than that of the NS groups sharing similar clinical and acute care utilization profiles. Nevertheless, the findings have also indicated that non-specialist care such as nursing transitional care or primary care in the form of general outpatient clinic visit should also be regarded as high importance to 28-day rehospitalization outcomes. It therefore stands to reason that the different types of specialists and non-specialist care, even non-medical community care, could be coordinated to address the care needs of discharged patients in an integrated manner to alleviate the dependency on specialist outpatient follow-up services alone. After all, an effective primary health care system requires the integration of preventive, promotive, curative, rehabilitative, and palliative healthcare services (Behera et al, 2022).

4 B. Findings reflecting Objective 2

To drill down the population-level profiles and map the trajectory of the evolution of the prevention needs identified, studies have examined the different factors contributing to the time course of developing cancers among diabetes and the trajectory of COPD surviving in the community. Our analyses revealed that not only do biomarkers and diagnostic complications of metabolic dysfunction determine whether and the time course of diabetes patients developing cancers**: Yau et al, 2024a,** Scoring system for predicting the risk of liver cancer among diabetes patients: a random survival forest-guided approach)

Project: S2019.A4.015.19S

Liver cancer is the third leading cause of cancer death worldwide. However, no single risk scoring system is universally accepted, nor scoring focusing on non-viral etiologies (such as metabolic conditions) has been widely developed. This study aimed to develop a scoring system to identify high-risk diabetes patients in primary care settings by retrospective cohort study using territory-wide electronic health records of Hong Kong.

A total of 384,443 patients were initially identified. Among the overall cohort, the liver cancer incidence rates for females and males were 0.48 and 1.40 per 1000 person-years respectively. The corresponding rates for those who were aged 55 years or above in the absence of hepatitis B/C or cirrhosis (female: n=396; male: n=1,015) were 0.36 and 0.94 per 1000 person-years. Patients who developed liver cancer (n=2,034) and a subset of patients who remained free of cancer (n=40,680) during follow-up (median: 6.2 years), were selected for model building.

In the final sex-specific scoring system, presence of chronic hepatitis B/C, liver cirrhosis, age, waist-to-hip ratio, and high-density lipoprotein cholesterol (HDL-C) (optimal range: 0.89 to 1.6 mmol/L for female; 0.79 to 1.4 mmol/L for male) were commonly included. Waist-to-hip ratio and HDL-C were also identified as important variables for liver cancer prediction among males in the absence of chronic hepatitis B/C or cirrhosis.

Among patients aged 55 years or above in the absence of chronic hepatitis B/C or liver cirrhosis, for males, waist-to-ratio and HDL cholesterol were consistently selected as variables in the scoring system accounting for almost 80% of risk in total in the scoring system. Waist-to-hip ratio ≥0.87 started to elevate the risk of developing liver cancer, however, the marginal increase of normalized risk between 0.87 and 0.90 (cut-off defined by WHO) among males without known liver diseases were more pronounced. Duration of diabetes ≥20 years, HbA_{1c} ≥8% and serum creatinine (optimal range was identified as 75 to 104μmol/L) were also selected. Age was no longer identified as important predictor of liver cancer risk in males. For females, age accounted for approximately half of the risk in the scoring system. While COPD and heart failure contributed to over 40% of risk in total, HbA_{1c} ≥7.8% and duration of diabetes ≥12 years accounted for the remaining.

Among the overall cohort on test set for model building, for females with score intervals ranging from 0 to 39, 40 to 49, and 50 or above, the 5-year liver cancer-free survival probability was 0.986, 0.944, and 0.897 respectively, while for males with score intervals ranging from 0 to 49, 50 to 69, and 70 or above, the 5-year survival probability was 0.968, 0.854, and 0.533 respectively. When the proposed scoring was applied to the entire cohort, for females and males with score interval 80 or above, the 5-year survival probability was 0.749 and 0.807.

Metabolic indicators such as waist-to-hip ratio and HDL cholesterol may also serve as markers for liver cancer development apart from age and preexisting liver diseases. These metabolic indicators remain important markers for liver cancer among male patients aged 55 years or above without known liver diseases. COPD, heart failure, and serum creatinine are also potential indicators of liver cancer for those without known liver diseases.

Another study "**Differential metabolic dysfunction profiles and site-specific risk of obesity-related cancers in patients with diabetes**" (Yau et al, 2024b) further examined risk of obesity on other cancers. A retrospective cohort study was conducted using territory wide electronic records and diabetes patients receiving care in public general outpatient clinics between 2010 and 2019 without history of malignancy were included to follow up until December 2019 and colorectal, liver, pancreatic bladder, kidney cancers were outcomes of interest.

Among patients with diabetes, central obesity was associated with an increased risk of colorectal, bladder, and liver cancers. Elevated glycaemic levels were linked to a higher risk of liver and pancreatic cancers. Serum lipid profiles have demonstrated mixed association with the risk of liver and pancreatic, kidney and gastric cancers. Renal dysfunction was also linked to an elevated risk of liver, bladder, and colorectal cancers.

The biomarkers and the diagnoses (when comorbid with COPD), also shorten the lifespan of COPD patients (Leung et al (2024b)**, Article- Comorbidities and prior medical service utilizations affect the mortality trajectories of Chronic Obstructive Pulmonary Disease (COPD) patients: Stratifying 2- and 5- year survival probability of 113,754 community-dwelling 65+ COPD patients discharged from Hong Kong public hospitals with a random survival forest-based decision tool**).

The model stratified 2- and 5-year mortality risk with the performance of AUC=0.81 and 0.80, respectively, and the patients' cumulative length of stay and frequency of hospitalization during the past year, as well as one's comorbidity-associated multiplications in the levels of resources consumed at index hospitalisation and histories of chronic illnesses such as lung cancers and heart diseases, were of top marginal importance to stratifying time-to mortality over a 2- or 5-year span even after adjusting for different biomarkers for lung functions and general clinical acuity at index hospitalization, such as urea, creatinine, GOLD classification (which was not selected by the model to stratified 2- or 5-year risk due to low marginal importance to outcomes). Historical medical service utilization and comorbidities, both at index hospitalization and historical, out-ranked the more frequently studied biomarkers for lung functions and general clinical acuity at index hospitalization in predicting the survival of COPD patients. The results underscore the need of better managing re-hospitalisation risk and chronic illnesses among COPD patients in the community during one's post-discharge period.

4C. Findings reflecting Objective 3

Two community cohorts were recruited from the respect districts and examined with profiles that emerged from the EHRs were also observed, and if the inclusion of features missing from the EHRs, such as features of the functional, psychosocial, demographic, and quality of life nature will enable additional factors that predisposing prevention needs over the course of the study. Data mining performed on the two cohorts revealed that, on the one hand, pain and how pain interfered daily activities and mental health emerged as a viable topic in K&T, whereby biomarkers of metabolic dysfunction predicted the interference of pain in daily activities and mental health in the K&T cohort (Leung et al, 2024c, Published Article- Impact of environment on pain among the working poor: Making use of random forest-based stratification tool to study the socioecology of pain interference. *International Journal of Environmental Research and Public Health* 2024 *21*, 179. https://doi.org/10.3390/ijerph21020179). The finding can optimize the cost-effectiveness of the service mix offered at the District Health Centers of K&T and SSP, which may enable the operators of K&T and SSP District Health Centers to better integration between medical and social services of K&T and SSP

The random forest algorithm was utilised to model and risk-score the unique and combined contribution of a diverse ensemble of environmental, sociodemographic, and clinical factors to pain interference among the working poor. The model is a commonly used machine learning algorithm to combine the output of multiple decision trees to enhance their performance, lower the requirement of sample sizes, while maintaining interpretability of the findings. A risk-scoring system was subsequently developed from the result of the random forest model with validation.

Pain interferes with one's work and social lives and, at the personal levels, daily activities, mood, and sleep quality. Little research has been done on pain interference and its socioecological determinants among the working poor. A novel random forest algorithm was deployed to model and quantify the unique contribution of a diverse ensemble of environmental, sociodemographic, and clinical factors to pain interference meeting the clinical/policy decision needs, and the technical challenges of isolating the intricately interrelated socioecological factors contributing to pain interference. The model can also help quantifying the relative contributions of each factor in an interpretable manner to inform clinical and policy decision-making.

The individual-level random forest model has revealed that age, occupation, the intensity and location of pain, BMI and blood glucose of those suffering from pain were found consistently the top-ten features contributing to

pain interference across all areas studied. When the model applied to a feature pool consisting of both individual- and building-level features, the random forest models supervised by mood, daily activities, sleeping quality, social life, and work performance yielded AUCs=.96, .98, .94, .99, and .99, respectively. Features such as age, occupation, pain intensity and BMI were found to remain in the top ten in terms of their summative scores. Comparing to those received from individual-level-only models, the cardinal orientation of the building, the proportion of flats with less than three non-functional rooms, the building age , and the number of corridors and lifts were found to be the top-ten importance according to the summative scores and for most of the individual studied areas of pain interference.

The analysis of this study has revealed that features representing internal built environment of working poor, such as the size of the living space, air quality, access to light, architectural design conducive for social connection, and age of the building, were assigned greater statistical importance than other more commonly examined predisposing factors for pain interference such as age, occupation, the severity and locations of pain, BMI, serum blood sugar, and blood pressure.

Selected items from the assessment of activities of daily living and instrumental activities of daily living were found consistently top of the list in predicting frequent hospitalization and the effects were robust before COVID-19 and were also robust at the peak of COVID-19 (Refer to Leung et al 2024d **Screening for Frequent Hospitalization Risk among Community-dwelling Elderly between 2016 and 2023: Machine learning-driven item selection, scoring system development, and prospective validation)**

The following items from the full CGA were selected by the model showing the best performance:

- presence of polypharmacy/the number of medications;
- presence of ADL issues such as requiring assistance with emptying bowel and bladder;
- presence of ADL issues such as requiring assistance with housekeeping, transportation, laundry, medication and shopping, food, and finances;
- presence of issues with mobilities;
- and the presence of chronic illnesses such as COPD, heart diseases, and depressive symptomology.

The data extracted from social service systems (**Article- Leung et al (2024f). The Valuation of Elderly Homecare Services Under a Joint Medical-Social Budget Perspective**) and unstructured medical notes put together by physicians (**Article- The presence of living arrangements and activities of daily living as topics in 257,470 medical notes of palliative patients with unsupervised Natural Language Processing (NLP) and random forest models** have demonstrated that services that provide assistance to instrumental activities of daily living is 1) most cost-effective for the long term irrespective of the functional and deterioration profile of the patients and what other services the clients were receiving, 2) slow down the entry into the most resource-intensive period of an elderly life, the end-of-life period, and thus incurred cost saving, and 3) essential to assess the palliative patients' needs for, otherwise will hasten mortality when discharged.

Instrumental homecare, on its own or combined with either one or both of the other home care services, has been shown to yield the greatest costs savings compared to other services. When expressed under a joint medical-social budget perspectives, instrumental home care would reduce medical costs of HK$34.53 and HK$85.03 for every dollar invested in instrumental-restorative home care respectively.

The study of NLP revealed that the topics of living arrangements and/or ADL were found to be most likely among high-need palliative patients. After controlling for patients' clinical and utilisation profiles, medical notes of palliative patients included the topic of living arrangements and ADL showed a 13% greater 90-day survival probability than palliative patients whose medical notes did not include those topics.

And as committed in the proposal, algorithms for optimizing the mix of services according to cost-effectiveness under a joint medical-social budgetary framework have been provided by **published article "Robust meal delivery service for the elderly: a case study in Hong Kong Scientific Report 2022; 12 (**https://doi.org/10.1038/s41598-022-25924-6 **).**

4D Findings reflecting Objective 4

Finally, with COVID-19 having emerged shortly after the commencement of the current study as being the most immanent primary prevention targets, and preventing its associated hospitalization the most critical secondary one, hence, the three studies were conducted to profile primary and secondary prevention needs of the studied districts with respect to COVID-19 and to map the trajectory of how primary and secondary needs associated with COVID-19 evolved for the forecasting of the growing demands of services, and one study was conducted to the sociodemographic and environmental effects of surge of percentage of ED waiting time exceeding 4 hours during the period between waves four and five.

- Guan, Leung, Kwok et al (2023): "Quantifying the Risk of General Health and Early COVID-19 Spread in Residential Buildings with Deep Learning and Expert-augmented Machine Learning". *10.2139/ssrn.4432448*)),
- Guan et al (2024) : "Applying Deep Learning and Expert-augmented Machine Learning to Quantify the Risk of Poor General Health and COVID-19 in Public Housing Residents' Built Environment: The role of socioecology during the first 365 days of COVID-19.
- Leung et al (2024g) on Article on "Deep Learning Approach to Forcasting COVID-19 Cases in Residential Buildings' of Hong Kong: A Multi-soueces Data Analysis of the Differential Roles of Environments and Sociodemographic during the Emergence and Resurgence of the Pandemic. *ArXiv* https://arvix.org/abs/2403.15759
- Leung et al (2024h) Article "Analysing the Variations in Emergency Department Boarding and Testing the Transferability of Forecasting Models across COVID-19 pandemic Waves in Hong Kong: Hybrid CNN-LSTM approach to quantitfying building-level sociological risk *ArXiv*: 2403.13842(cs) (http://arvxiv/abs/2403.13842).

The current study on quantifying the risk of general health profile and early COVID-19 spread in residential building by Gaun et al (2023) developed a deep-learning approach with multiple input channels to capture the hierarchical relationships among an individual's socioecology's demographical, medical, behavioural, psychosocial, and built-environment levels. The findings supported:

1) deep-learning models whose inputs were structured according to the hierarchy of one's socioecology outperformed plain models with one-layered input in predicting one's general health outcomes, with the model whose hierarchically structured input layers included one's built environment performed best;

2) built-environment features were more important to general health compared to features of one's sociodemographic and their health-related quality of life, behaviours, and service utilization;

3) a composite score representing built-environment features' statistical importance to general health significantly predicted building-level COVID-19 case counts; and

4) building configurations derived from the expert-augmented learning of granular built-environment features of high importance to the general health were also linked to building-level COVID-19 case counts of external samples.

The statistical importance of each granular internal built environment feature studied in each of the 35 buildings in District A has been shown by Poisson regression revealing that the composite score representing the aggregated internal built-environment features' importance to general health outcomes was significantly associated with first-365-day cumulative COVID-19 case counts of the same buildings with an incidence rate ratio of 5.83 (95% CI: 3.03-10.91; p-value <.001). The statistical importance of building configurations derived from the expert-augmented learning of granular built environment features of high importance to the general health of residents

of District A's buildings were mapped onto the studied buildings in District B. The three modular features of building configuration were:
1) being connected to other buildings,
2) being part of a building block, and ;
3) construction types.

Poisson regression revealed that the composite score representing the aggregation of the feature importance of the three modular features of building configuration identified from District A's buildings studied predicted the cumulative COVID-19 case counts of of the 18 buildings in District B's most prominent public housing estate onto which the modular configurational features could be mapped (incidence rate ratio=450.02 (95% CI: 24.47-17913.29; p-value <.001).

The study by Guan et al (2024) developed a deep learning model with multiple input channels to capture the hierarchical relationships among the demographical, medical, behavioural, psychosocial, and built environment levels of an individual's socioecology. The findings supported the hypotheses:

(1) deep learning models whose inputs were structured according to the hierarchy of one's socioecology outperformed traditional models with one-layered input in predicting one's general health outcomes, with the model whose hierarchically structured input layers included one's built environment performed best;

(2) built environment features were assigned greater statistical importance to general health compared to features associated with one's socio-demographics, health, and health-related behaviours and service utilization;

(3) buildings' built environment features' statistical importance to general health outcome, when aggregated into a composite score, significantly predicted their cumulative COVID-19 case counts; and

(4) building configurations derived from the expert-augmented learning of granular built environment features that were of high importance to the general health in one district were also linked to COVID-19 case count accumulated in buildings of another district.

This study adds further evidence that the built environment that put residents at risk for poor general health also put them at risk for COVID-19.

Study by Leung et al 2024g on "Residential Buildings' Internal and External Environments and Residents' Sociodemographic Profiles Differentially Predicted Buildings' COVID-19 Cases Accumulated during the Early Phases vs. the Resurgence of the Pandemic" applied a multi-headed hierarchical convolutional neural network to structure the vast array of hierarchically related predictive features representing buildings' internal and external built environment and residents' sociodemographic profiles to model COVID-19 cases accumulated in buildings across three adjacent districts in Hong Kong, both before and during the recent resurgence of the pandemic.

The model shows that different sets of factors were found to be linked to the earlier waves of COVID-19 outbreaks compared to the recent resurgence of the pandemic. The sociodemographic factors such as work hours, monthly household income, employment types, and the number of non-working adults or children in household populations were of high importance to the studied buildings' COVID-19 case counts during the early waves of the pandemic. Factors constituting one's internal built environment, such as the number of distinct households in the buildings, the number of distinct households per floor, and the number of floors, corridors, and lifts, had the greatest unique contributions to the building-level COVID-19 case counts during the resurgence.

The ED study aimed to forecast breaches of the 4-hour ED waiting time target between waves four and five, when the pandemic was subsiding and coexisting with the virus (Leung et al, 2024h. Data were extracted from multiple

public sources. The Hospital Authority provided hourly records of the emergency department (ED) waiting time for the studied hospital, while the Department of Health made available the accumulated COVID-19 case counts for each building in the studied district throughout the pandemic. ED waiting time records were extracted for the period between December 31, 2018, and July 27, 2022. Daily COVID-19 confirmed cases in buildings within the catchment area of the hospital were extracted separately for Wave One to Four (January 23, 2020, to May 21, 2021), the period between Waves Four and Five (May 22, 2021, to December 23, 2021), and Wave Five (December 24, 2021, to July 23, 2022). Additionally, data related to the sociodemographic characteristics of residential buildings, as well as their internal (architectural) and external built environment, were obtained from the Census and Statistics Department, Housing Authority, and Google Map, respectively.

The result shows that the historical surge pattern, case counts, and factors associated with a residential built environment and sociodemographic profiles were found to be critical for forecasting breaches of the 4-hour target during different phases of COVID-19. Historical surge patterns and case counts were the key contributors to forecasting ED's surges over a 14-day horizon while factors associated with the residential built environment and sociodemographic profiles were critical to forecasting surges between wave four and five epidemic. Additionally, transfer learning between waves four and five enhanced the learning of surge patterns in waves one to four and the wave five.

The main driver for surges in ED attendance during the period between waves four and five are the adjusted effects of sociodemographic and environmental factors, which are stable over time comparing to historical surge patterns and cases counts that changes daily and across the pandemic, making this period be potentially informative for preparing our medical system for future endemic outbreak not only because the current phrase of the pandemic and the period of the brief reprieve from the virus after wave four were both characterised by “living with the virus.”

Chapter 5: DISCUSSION & POLICY RECOMMENDATIONS

Overall view

Highlighted below are discussions and recommendations generated from findings resulted from the multi-source data analysis performed under the CDC's socio-ecological framework for prevention. Hence, the findings were presented below in accordance with the spheres laid out in the framework and the study objectives. The artificial intelligence algorithm can map the trajectory of how primary, secondary and tertiary prevention needs to forecast the growing demands of medical and social services of the population. The "one-size-fits-all" approach to multidisciplinary care is not sufficient in addressing the needs of patients with multiple and chronic health conditions, who may encounter many healthcare professionals from different disciplines across primary and secondary care settings. A brief overview of key findings and recommendations will first be presented followed by more detail discussions and policy implications under different studies.

In terms of tertiary-wide policy recommendation, our analyses revealed a high degree of heterogeneity among residents of different HK districts in terms of their medical and social care needs, their built environment and sociodemographic's impact on their health, and their health- and healthcare-related behaviours responding to health and social policy (including COVID-19-related public health and social measures). It is our observation that the District Officers' unique perspectives on their respective constituents' environment, behavioral patterns, and care needs can benefit the government's development and implementation of health and social policies. Therefore, we recommend that not only should the District Officers be involved in the government's development and implementation of health and social policies, but we also recommend that that additional investments be made to ensure that the District Officers are well-informed about the residents in their district, by providing access to appropriate data and data visualization/decision-support tools.

At the level of a district's built environment, our analyses revealed that specific configurations of residential buildings' internal and external environment put their residents at higher risk for poor general health, acute hospital admission, and COVID-19 infections. Hence, we recommend a geographical data-driven approach to screening for residents who can benefit more from primary and secondary prevention by targeting residential buildings that fit specific high-risk built-environment profile. In addition, we recommend that resident's access to gyms and recreational facilities should be enabled, as our analyses revealed that such access can mitigate the potentially high acute medical care needs and COVID-19 infection risk predisposed by specific architectural elements and sociodemographic profiles of public housing.

In terms of the districts' service ecology, different medical and social service boundaries are drawn within a district, and its residents who dwell within the catchment areas are entitled to be served by these services. Not only had our analysis revealed that social services and ambulatory and postacute services reduce re-hospitalization and are therefore effective means of tertiary prevention, but it had also demonstrated that, with data-driven optimization programming, the resultant optimal pairing among care needs of the service end-users, service types, and the timing of service provision can maximize service efficiency and medical cost savings for the district's health and care system. Hence, we recommend that the government to take a joint medical-social budgetary perspective and evaluate the cost-effectiveness of social services and ambulatory and postacute services in terms of the amount of medical utilization cost they can potentially save.

At the person-level, individual residents' trajectories of care needs (also reflected in the corresponding utilization over time), as our analyses had shown, were not only determined by their medical diagnoses but were also shared by their degrees of dependency associated with (instrumental) activities of daily living in a predictable manner. While the mechanism to qualify older adults who lack functional independence exists in HK, and HK's rate of

institutionalizing older adults exceeds that of other regions, our analyses of different community samples show that older adults with low functional dependency remain highly prevalent in the studied communities. The high prevalence of functionally dependent older adults in the community without proper levels of care to address their lack of independence in (instrument) activities of daily living could be attributed to the long queues for admitting those judged as lacking functional independence into institutional care or other types of long-term care.

The high prevalence of functionally dependent elderly may be associated with the widening of the gap between the amount that our government invested in community services for addressing the functional needs of older adults over time and the growth of the elderly population and the government's continued diversion of investment to services for frail elderly away from services that address the functional needs of non-frail elderly. It is of note not only because unmet functional needs linked were to frequent hospitalization and accelerated institutionalization of older adults in our studies, but for those elderly entering the period of greatest utilization of medical services, i.e. the end-of-life period according to the literature, unmet functional needs also shortened their lifespan – as demonstrated by our natural language processing of unstructured textual physician notes for recording patient care needs and communicating the needs to fellow physicians and other health professionals on the care team. It is, therefore, our recommendation that greater investment be made to screen for those who lack functional independence in the communities and provide better support to District Officers' leadership over the Care Team as a vehicle for monitoring the changing statuses of elderly residents' independence in the activities of daily living in different communities. In addition, to overcome challenges of government-commissioned social services' saturated capacity, it is also our recommendation that well-regulated public-private partnerships be foraged to entice social enterprises and the private sector to enter into the market to bridge the current service gaps and address the needs arising from the lack of independency in (instrumental) activities of daily living.

Our findings on the long-term generalized cost-effectiveness of social services under a joint medical-social budget framework may inform the trade-off between market incentives and social goods in regulatory decisions. In addition to being vehicle for monitoring changes in elderly residents' independence in the activities of daily living, the Care Team can also be instrumental in screening residents for other determinants of medical and social service utilization over time that our different databases have consistently shown to be of relevance. For example, metabolic dysfunction has also been shown to be contributing to the emergence of prevention needs of the studied districts and shaping the trajectory of corresponding service utilizations. Analyses of data collected from the community cohort demonstrated that those who suffered from metabolic dysfunction were also more likely to suffer concurrently from psychosocial issues, functional dependency and comorbiding chronic illnesses. In addition, with population-level analyses, we have also shown that behaviours that contribute to metabolic dysfunctions, such as smoking and poor metabolic control with medication, are also linked to diabetes and the development among diabetes the cancer types that were typical of high-prevention-needs profiles of K&T and SSP residents. Hence, it is also our recommendation that the Care Team be monitoring the metabolic dysfunction and psychosocial well-being of the residents of studied districts, especially those who suffered from diabetes, and supporting behavioural modification programs targeting, for example, smoking cessation and exercises. Behaviour modification programs on smoking cessation and exercises can also serve as the basis for community rehabilitation programs for COPD patients, whose trajectory of resource utilization is pervasive and intensive as shown in our population-level studies, and their tertiary prevention needs were consistently the highest of all patient profiles identified from the K&T and SSP populations (see below section). It is therefore our recommendation that the Care Team can be trained as helper to be involved in community rehabilitation programs for COPD patients to support the work of professionals.

When profiling attributes of residents, in addition to ones' functional and psychosocial health and behavioural health, our analyse have also revealed that tertiary preventions needs (in terms of the likelihood for 28-day re-hospitalization) were greater for K&T and SSP residents who received diagnoses such as COPD (for young-old and elderly patients) and digestive cancers (young-olds only) - even compared to patients whose diagnoses belonged to the same major clinical categories of diagnoses and shared similar clinical acuity and complexity.

This is of great policy implications as, first of all, COPD is considered an ambulatory care sensitive condition, which does not require acute medical care if it were managed appropriately in community and primary care settings. Hence, the fact that COPD, an ambulatory care sensitive condition, is not only deemed a "typical" patient profile for both K&T and SSP, but COPD is also the diagnosis of the "typical" patients who re-hospitalization within 28-days of discharge may suggest that we can invest more tertiary prevention effort into patients with COPD. As it stands, COPD patients is not among the target groups that the DHCs are mandated to serve.

It is our recommendation that DHCs and other primary care initiatives be taking into consideration COPD patients when planning their tertiary prevention strategies. We also recommend that better support be provided to the District Officer to facilitate the management of the Care Team to engage in non-professional-led self-management and community rehabilitation of COPD patients. Secondly this is also of great policy relevance that young-old patients diagnosed with digestive cancers were the "typical" profile for the K&T and SSP patient populations who re-hospitalized within 28 days of discharge.

Unlike their 65+ counterpart, 50-64 patients with digestive patients lack sufficient community services and a continuity of care services in the community to which they were discharged. It is our recommendation that more resources should be invested into community services catering to the needs of digestive patients aged 50-64, many of whom are working and taking care of their offspring and parents at the same time. In fact, for young olds in the K&T and SSP inpatient populations, those who re-hospitalized were more likely the ones who did not received post-discharge specialized outpatient follow-up care due to their being quick to re-hospitalize (for example, median time to re-hospitalization for 50-64 is less than 2 weeks while SOPC appointment time falls between 9 weeks to 111 weeks depending on the specialties (https://www.ha.org.hk/haho/ho/sopc/dw_wait_ls.pdf)). Our analyses have also shown that those 50-64 patients who re-hospitalize before they could receive SOPC or other post-acute services from HA were also significantly more likely to be suffering from two or more chronic illness compared to their 65+ counterpart. It is therefore our recommendation that government introduced better incentive to entice a greater number and variety of timely post-discharge services for young-old patients discharged into the community and to enhance the self-management of chronic illnesses among young olds in K&T and SSP.

Study: Driving precision public health research and intervention in Hong Kong (Leung et al, 2024)

Discussion of findings

This study with population data seeks to contribute to patient segmentation research by going beyond the dominant approach in the literature that focuses on identifying sub-populations solely based on diagnostic homogeneity (Chong et al, 2019; Mechanic et al 2014; Nnohaam et al, 2020; Lafortune et al 2009; Liu et al 2012; Eisserns et al, 2014; Joynrt et al, 2017) or costs (Davis et al, 2018). Instead, this study's partition of the population is optimised to reveal the population's unmet care needs and the corresponding service gaps. The findings highlighted three reasons why such care integration for tertiary prevention is of particular relevance to the healthcare system of the studied districts.

Firstly, the diagnoses of typical patients of the segments of the inpatient population of 50-64 and 65+ are angina and arrhythmia, respectively, and the segments' respective NS groups consist of typical patients diagnosed with heart failure. Angina and arrhythmia would lead to heart failure if not well managed. Nevertheless, angina and arrhythmia, as well as heart failure, could benefit from cardiac rehabilitation in the community setting to reduce rehospitalisation and mitigate deterioration into heart failure. Managing medical risk factors with cardio-protective therapeutics alone is insufficient, and psychosocial care and lifestyle management are also critical (Dala et al, 2015). Similarly, COPD was diagnosed among 50-64 and 65+ patients typical of the inpatient populations studied and the NS group of a segment of the studied population. Although COPD is one of the leading causes of readmission to acute care hospitals worldwide, research has shown that ongoing clinical and nursing support alone is insufficient. A multidisciplinary pulmonary rehabilitation (PR) program making available

health-related information, advice on exercise programs, addresses cognitive and behavioural issues, and tailors care plans to individual's needs is most effective in reducing re-hospitalization of COPD patients (Bourbeau et al, 2003; Cravo et al, 2022).

Secondly, chronic illnesses such as cardiovascular diseases and COPD were found to be more prevalent in the NS group and were associated with higher 28-day rehospitalization rates, but chronic illnesses were also more likely to co-occur with one another among the NS groups. In addition, NS groups also reported greater resource-intensifying comorbidity during index acute care hospitalisation. Research has shown that the traditional '*assess-and-advise*' model of primary care is not sufficient for chronically ill patients especially when they also have multiple health problems. Instead, comprehensive and holistic care with good coordination is essential to help these patients to navigate the complexity of the healthcare system (Lee, 2020). The study found that the NS groups of 50-64 were consistently reporting greater resource-intensifying comorbidity at index hospitalisation, greater prevalence of having two or more chronic illnesses historically, and higher 28-day rehospitalisation rates compared their 65+ counterpart. The findings are consistent with recent studies demonstrating that younger patients of diabetes showed a greater relationship between the number of co-morbidities, on the one hand, and increased mortality and healthcare costs on the other (Hong et al, 2023).

The outcomes of prevention programs for younger patients were noted to be better across a wide range of targets, especially smoking cessation, compared to elderly patients (Al Quait and Doherty, 2016). Of particular relevance to the studied population of 50-64 is that the smoking predisposes many of the clinical conditions that characterise the typical patients of the population or its NS groups, such as COPD, heart failure, malignant neoplasm of urinary system, or digestive malignancy.

In addition to being more effective than programs targeting elderly patients across many targets, programs targeting younger patients also differ from ones targeting elderly in terms of their focus on empowering active coping, such as in the case of young patients who experienced severe pain (Soares et al, 2004) which is of relevance as the typical patients of the 50-64 population were those who suffer from inflammatory and reactive arthropathy. Effective primary care tertiary prevention programs of smoke cessation and pain management is a multidisciplinary approach to enhance self-management among patients suffering from chronic conditions (Lorig, 2006).

Thirdly, the analyses revealed that the typical patients identified from almost all of the full segments and some of the no-service sub-populations were diagnosed with ambulatory care-sensitive conditions (ACSC) as defined by Lin et al (2017), "ACSC inpatient admissions as potentially avoidable if timely and effective care had been received in an ambulatory care setting." It has been noted that redistributing the acute care resources currently consumed by ACSC patients to manage their long-term conditions through interdisciplinary primary care can prevent avoidable hospitalization (Grabowski and Mor, 2020).

Policy implications

The finding of this particular study advocates for more innovative approach taken by the community to address the care gaps of the discharged patients who are at elevated risk for re-hospitalisation within a shorter timeframe. For example, given that multimorbidity among dementia and other neurological and mental health patients contributes to high hospitalisation rates due to ACSCs, access to proactive multidisciplinary disease management programs could effectively reduce ACSC-related hospitalisations and readmissions among dementia and other neurological and mental health patients (Wolf et al, 2019).

Here, care needs and service gaps were identified from the studied in-patient population using a precision public health approach that is augmented by learning algorithms (Leung et al, 2023). As the needs of our ageing population grow in intensity and diversity, there is a need to bring precision to the current multidisciplinary approach to preventive care. A protocol-based "one-size-fits-all" approach to multidisciplinary care is not

sufficient in addressing the needs of patients with multiple and chronic health conditions, who may encounter many healthcare professionals from different disciplines across primary and secondary care settings.

If a personalised care approach is lacking, patients receiving multidisciplinary care may find themselves recounting their life histories in every new encounter. It has been recently demonstrated in the care of COPD patients that a case management approach to personalised multidisciplinary care can best facilitate a trans-disciplinary approach to healthcare across the patient journey (Carvo et al, 2022), with healthcare professionals that meet the most pressing needs of that particular patient at the initial stage taking up the case manager role (Lee, 2014). Case manager should be supported by a community health practitioners (a junior health practitioner with general health training without status of particular health disciplines) as case-coordinator to help the case manager to co-ordinate other professional services for the patient (Lee, 2014). The case co-ordinator can support the case management once the condition is established as the main task will then be enforcement of compliance to professional advice and mobilisation of resources and support the patient to be managed in community. This will allow the professionals to care for new cases.

Study: Scoring system for predicting the risk of liver cancer among diabetes patients: a random survival forest-guided approach (Yau et al, 2024a)

Discussion of findings

Findings of this study identified waist-to-hip ratio as a predictor of liver cancer among the overall cohort and older males in the absence of chronic hepatitis B/C or liver cirrhosis. While obesity (mainly using BMI as indicator) has been linked to liver cancer (Lauby-Scretan et al, 2016),[3] obesity indicators have been rarely included in scoring systems for viral etiologies (Kubota et al, 2020). BMI was included in both Cox regression and random forest models based on a priori knowledge (Singel et al, 2013). BMI was not identified as risk factor for liver cancer among diabetes patients (An et al, 2021; Franczyk et al, 2021). Results of this study may suggest that among the general population in the presence or absence of existing liver diseases, other obesity indicators (such as waist-to-hip ratio which better reflects body fat distribution), not just BMI alone, may also be considered as candidate variables for liver cancer scoring. Evaluation of whether a lower cutoff of waist-to-hip ratio for Asian males to define abdominal obesity would be beneficial to health outcomes is warranted.

Findings of this study may imply that HDL cholesterol above 1.6 or 1.4 mmol/L for female and male respectively may not necessarily confer additional benefits against liver cancer risk or potentially overall health risk. Similarly, HDL cholesterol below the widely accepted threshold may not increase liver cancer risk until it drops below 0.89 or 0.79 mmol/L for female and male respectively. The optimal range identified against liver cancer risk in this study appears to be consistent with the findings of a previous study (Allard-Ratick et al, 2018),[24] which demonstrated that HDL cholesterol below 0.78 or above 1.55 mmol/L may elevate the risk of cardiovascular events.

Results of the current study has shown that among patients without known liver diseases, variables incorporated in the liver cancer risk scoring for male could be more easily modifiable than those for female. This study also identified several important variables in scoring uniquely to those without known liver diseases, for example, COPD and heart failure for female, and serum creatinine for male. While evidence linking COPD, heart failure, or renal dysfunction to liver cancer remains limited in the literature, in a recent study conducted in Korea, Ahn et al (2020) demonstrated that COPD is not only associated with lung cancer, but also liver cancer. On the other hand, in a recent review, Bertero et al (2018) summarised that the observed higher incidence of cancer among patients with heart failure than those without in epidemiological studies could be due to shared risk factors between cardiovascular disease and cancer. Furthermore, serum creatinine either too low or too high could be indicative of elevated liver cancer risk. While low serum creatinine may potentially signal an undiagnosed liver disease leading to reduced hepatic production of creatinine (Slack et al, 2010), high serum creatinine may indicate

underlying renal dysfunction, which may also be patho-histologically linked to liver cancer through accumulation of carcinogenic toxins, oxidative stress, and chronic inflammation (Shi et al, 2023).

Policy implications

This study has enlightened us that metabolic indicators such as waist-to-hip ratio and HDL cholesterol may also serve as markers for liver cancer development apart from age and preexisting liver diseases. These metabolic indicators remain important markers for liver cancer among male patients aged 55 years or above without known liver diseases. COPD, heart failure, and serum creatinine are also potential indicators of liver cancer for those without known liver cancer.

The findings have significant implications not only on prevention of liver cancer and also prevention of further complications of people with other common chronic illnesses to develop liver cancer such as not well controlled diabetes mellitus, COPD, heart failure, chronic kidney disease (those with raised creatinine) even without known liver diseases. Universal vaccination of Hepatitis B may not concur benefits among elderly as they did not receive the vaccination nor lower level of protection if vaccinated at older age. However, they can be at higher risk of developing liver cancer particularly female. Identifying those with chronic illnesses as mentioned early on would serve as secondary prevention of liver cancer (identification of predicting risk factors for liver cancer) and also tertiary prevention for those chronic illnesses (preventing development of complications such as liver cancer). The findings provide added evidence the importance of early screening of those with potential risks of developing hypertension and diabetes mellitus which can link to high risk of development of chronic illnesses. The Chronic Disease Co-Care (CDCC) pilot scheme initiated by Primary Care Commission must be firmly in place for better management not only for hypertension and diabetes to minimise development of cardiovascular and renal complications and also liver cancer, the leading cancer death in Hong Kong and globally.

Primary prevention to promote healthy eating and regular physical activities preventing obesity (waist-hip ratio) and optimal lipid level can also facilitate cancer prevention such as liver cancer. The machine learning-guided approach adopted by this study can have the potential to develop other cancer risk scoring systems among patients with chronic illnesses such as diabetes mellitus, hypertension and also possession of metabolic and cardio-vascular risks. The findings of this study demonstrates that artificial intelligence algorithm can map the trajectory of how primary, secondary and tertiary prevention needs to forecast the growing demands of medical and social services of the population. One must not forget the importance of primary prevention on smoking as smoking related diseases such as COPD, heart diseases also putting people at higher risk of liver cancer.

The findings presented here showed that factors precipitating COPD patients' mortality outcomes over longer timeframes are potentially different from factors associated with a narrower window of observation examined in previous studies. Notably, while previously examined factors for mortality outcomes observed over a shorter timeframe were primarily biomarkers of lung functions and general clinical acuity that could only be ascertained via the medical laboratory, factors precipitating COPD patients' mortality outcomes over a longer period are information on utilization and comorbidities that could potentially be ascertained via self-report at or shortly after being discharged from index hospitalization. Hence, the findings presented here are therefore of relevance to informing the integration between medical and social care and the planning of community-based end-of-life care for COPD patients.

Study on Differential metabolic dysfunction profiles and site-specific risk of obesity-related cancers in patients with diabetes (Yau et al, 2024b)

This study has further examined the differential contribution of a comprehensive profile of metabolic dysfunction-related biomarkers, lifestyle factors and co-morbid chronic illnesses to different cancer outcomes among the diabetic population, The identified key parameters such as abdominal obesity, glycaemic control and lipid profile can serve as cancer prevention targets among diabetes patients. Regular monitoring for diabetic

patients may also help to improve cancer outcomes so CDCC can play an important role for potential cancer prevention targeting diabetic patients.

Study on Comorbidities and prior medical service utilizations affect the mortality trajectories of Chronic Obstructive Pulmonary Disease (COPD) patients: Stratifying 2- and 5-year survival probability of 113,754 community-dwelling 65+ COPD patients discharged from Hong Kong public hospitals with a random survival forest-based decision tool (Leung et al 2024b)

Discussion of findings

This study applied a machine learning algorithm that takes into the complexity among interrelated biomarkers for lung functions and general clinical acuity, as well as historical and contemptuous utilization and comorbidities, to model the survival outcomes over different timeframes and assign statistical importance according to each feature's unique statistical contribution to survival outcomes.

Factors precipitating COPD patients' mortality outcomes over longer timeframes are potentially different from factors associated with a narrower window of observation examined in previous studies which examined factors for mortality outcomes observed over a shorter timeframe primarily biomarkers of lung functions and general clinical acuity. Those could only be ascertained via the medical laboratory, and factors precipitating COPD patients' mortality outcomes over a longer period would be the information on utilisation and comorbidities. The information could potentially be ascertained via self-report at or shortly after being discharged from index hospitalization. The findings of this study are relevance to informing the integration between medical and social care and the planning of community-based end-of-life care for COPD patients.

Policy implications

Findings of this study further emphasise the need of better managing re-hospitalisation risk and chronic illnesses among COPD patients in the community during one's post-discharge period. The time-to-morality stratification tools of this current study can be administrated at discharge from index hospitalisation to guild interprofessional collaboration with primary care and community care providers in targeting chronic disease management and preventing re-hospitalization for improving the quality and longevity of the lives of COPD patients in the community.

CDCC scheme should consider including COPD with more structured community-based rehabilitation (CBR) programme to shorten length of stay in hospital and better management of co-morbid health conditions as well as pulmonary rehabilitation programme in community setting. CBR for COPD usually poses challenge as COPD patients tend to be re-admitted very soon after discharge. This creates a vicious cycle as prolong length of stay leading to higher mortality but delayed admission would also pose danger of life. Intensive CBR can also include better management of other co-morbid conditions as well as improvement of lung function. This would minimise re-hospitalisation.

Similar to the study of precision public health (Leung et al 2024a), effective care should be available in community setting to manage different long-term conditions through interdisciplinary primary care can prevent avoidable hospitalisation (Grabowski and Mor, 2020). Case management approach to personalized multidisciplinary care can best facilitate a trans-disciplinary approach to healthcare across the patient journey for COPD (Carvo et al, 2022). Co-existing COPD among diabetic patients has also shown to increase risk of liver cancer by Yau et al study (2024a). The findings of this study further call for integrated care not only across different disciplines and also different chronic health conditions. Each discipline should not work in silo nor each chronic health condition should not be managed in silo.

Study on Leung et al, 2024c, Published Article-The environmental, sociodemographic, and clinical determinants of pain interference among the working poor: The development and validation of a random forest-based stratification tool on the socioecological impact.

Discussion of findings

This study has revealed that the working poor's internal built environment, such as the size of the living area (parameterized as the number of non-functional rooms in the flat), the building of residency's access to light, the quality of air, and the age of the building, were assigned as much (and sometimes higher, depending on the supervisory outcomes) statistical importance than other more commonly examined predisposing factors for pain interference such as age, occupation, the severity and locations of pain, BMI, serum blood sugar, and blood pressure.

Our model assigned to selected build environment features associated with small living areas, low accessibility to light (cardinal orientation and the number of light well in the building), poor air quality (fewer number of light wells), high social connectivity (number of distinct household in terms of the number of flats, number of lifts, and the number of corridors), and older building are of high importance to determine whether the working poor might have suffered from pain interference, even after adjusting for unique and combined contributions of the clinical and demographic factors, as well as other built environment factors such as levels and size of the building Alternations in gray matter volume, white matter integrity and epigenetic changes in the brain were found in both pain patients and in rodent pain models by brain imaging studies (Bushnell et al, 2013, 2015) suggesting that these effects may possibly be prevented or reversed by environmental factors. Population data of this study can add evidence of the adverse effects of chronic pain and can be reduced or prevented by environmental and lifestyle factors.

The nearby nature would buffer the relation between catastrophising and pain intensity as nearby nature moderates the association between pain-related rumination and pain intensity, but not the helplessness-pain intensity or the magnification-pain intensity associations [Wells et al, 2019]. Healthcare professional might look for community resource such as nearby nature for pain management (Wells et al, 2019). A study in China has shown that people lived in moderate and unfavorable environments, had higher risks of arthritis in cross-sectional analysis and follow-up study so inferior living environment might promote the development of arthritis so improvement of the living environment can be good strategy for the primary prevention of arthritis (Liu et al, 2023).

The associations between built-environment quality and people's mental health were not found to be strong in Hong Kong (Kan et al, 2023). The impact of housing characteristics on mental health would be more direct in communities with relatively poor housing conditions, and the effect might be indirect for communities with relatively good housing conditions (Kan et al, 2023). It is more important to improve living environment of working poor and hope to improve their mental health to avoid aggravation their perception to pain.

Policy implications

There are many variables associated with health outcomes in real world of healthcare delivery and it is not also easy to predict the relationship between variables even with advanced statistics. Random forest algorithm is useful to for investigation of risk predictors of different health outcomes in the complex healthcare environment. This current study adds additional value to the literature with its novel application of the random forest algorithm to drive the development of a scoring system whose purpose is to identify the working poor at risk of pain interference across different areas in community settings.

There is benefit of informing community pain screening to target residential areas whose built environment contributed most to pain interference and informing the design of intervention programs to minimize pain interference among those suffering from chronic pain and showed specific characteristics. The findings of this

study can serve as a reminder for primary care physicians to obtain detail social history of their patients to identify those patients with high living environmental risks to assess pain and its interference for effective prevention. Early intervention would empower patients on pain management to minimise the pain and suffering to prevent further deterioration and restore the functional activity as far as possible. The findings support the call for good architecture to provide spirit and value of building.

Study on Screening for Frequent Hospitalization Risk among Community-dwelling Elderly between 2016 and 2023: Machine learning-driven item selection, scoring system development, and prospective validation. ***Frontier in Public Health*** **Forthcoming. Leung et al 2024d**

Discussion of findings

The implications of this research in the frequent hospitalization outcome-supervised item selection for the abbreviation of CGA demonstrates the improved accuracy of screening for "super-utilizers" (Vaida, 2017) in the medical system. The phenomenon of "super-utilizers" describes the 1% of the US population that accounted for 20% of its annual healthcare spending, and the 5% of the Medicaid beneficiaries responsible for more than half of its total expenditures (Adir et al, 2019). However, interventions offered to "super-utilisers" identified at inpatient discharge failed to demonstrate any significant impact in a large-scale and highly popularized randomized control trial (Finkelstein et al, 2020). Lack of demonstrable effectiveness of super-utilizer interventions has been attributed to the limited scope of items (primarily medical-related) being screened to inform the design of supposedly person-centre interventions (Lantz, 2020). Assessing items systemically to represent one's needs beyond medical ones is challenging in busy primary and acute care settings, especially when the number of non-medical factors precipitating one's frequent hospitalizations is too numerous to assess exhaustively (Adir et al, 2019]. In this study, a bottom-up data-driven approach has revealed that aspects of elderly population's functional dependency may amount to a greater unique contribution to frequent hospitalisation outcomes compared to the medical histories of elderly. The number of potential screening items can be drastically reduced to only ones with the highest outcome-differentiating powers and this study has also identified those items that could be screened for in the community setting.

Policy implications

This study has bridged the research gaps of lacking clinically relevant evidence for developing short-form CGA and out-performed published models whose objectives were to identify determinants of frequent hospitalization from isolated instruments selected from the CGA. The resulting short-form of CGA was developed into a scoring system that was valid in identifying frequent hospitalisation outcomes over time, despite COVID-19's impact on the communities, and across different geographical populations. Findings of this study has potentially provided a tool for community screening to prevent frequent hospitalization among the medical system's "super-utilisers."

Studies on Home care (Leung et al 2024f & Lin, Baker and Leung (2022) https://doi.org/10.1038/s41598-022-25924-6**)**

Policy implications

Instrumental homecare can increase hospitalization-free days among community-dwelling elderly and yield significant net system-wide cost savings. The study on “Valuation of Elderly Homecare services” (Leung et al 2024f) has demonstrated the feasibility of data-informed decision-making in system-wide resource allocation under a joint medical-social budget perspective.

The machine learning performed with random forest models of the study on palliative care (has revealed that the topics of living arrangements and/or ADL were most likely to be found among high-need palliative patients. Even after controlling for patients' clinical and utilization profiles, palliative patients whose medical notes included the topic of living arrangements and ADL showed a 13% greater 90-day survival probability than palliative patients whose medical notes did not include those topics. The study has shown how changes in government policy may have increased the relevance of the topic of living arrangements and ADL to physicians of palliative patients,

especially among those who showed greater clinical and utilisation needs, and those palliative patients whose medical notes include the topics of living arrangements and ADL showed a higher 90-day survival probability.

The study on "Robust meal delivery service for elderly" (Lin, Baker and Leung, 2022https://doi.org/10.1038/s41598-022-25924-6) has come up with some suggestions. Integration of multiple home care services into a single optimisation effort would be advantageous if different services, such as housekeeping, medical examination, and meal delivery, can be coordinated and optimized simultaneously. A van assisting workers in improving on-road transportation between different buildings (instead of walking) can be considered to improve the efficiency of the delivery system. The second direction is to investigate delivery economies where vehicles and staff can be tracked in real-time and planning decision epochs are much more dynamic and granular.

Considering a dynamically updated delivery system may better protect against real-time variable service times. As real-time decision-making largely depends on the internet of things (e.g., sensors) to obtain real-time information, designing a system that is remotely programmable to implement optimization algorithms within the sensor-based network can be an interesting direction. Another practical direction is to incorporate the elements of ride-hailing service into the meal-delivery system. Discussion on efficiency and fairness under such a new context would be promising to improve the elderly care.

Studies on profiling the primary and secondary prevention needs of the studied districts with respect to COVID-19

Discussion and policy implications

Guan et al study (2023) has shown that specific built environments would put residents at risk for poor general health and COVID-19 infections. The natural ventilation potential (NVP) is critical for airborne infection control in the micro-built environment, where infectious and susceptible people share air spaces. Taking Wuhan as the research area, the NVP in residential areas to combat COVID-19 during the outbreak was evaluated for four fundamental residential area layouts (point layout, parallel layout, center-around layout, and mixed layout) based on the semantic similarity model for point of interest (POI) picking (Lu et al, 2022). The analyses indicated that the center-around and point layout had a higher NVP, while the mixed and parallel layouts had a lower NVP in winter and spring. Further analysis showed that the proportion of the worst NVP has been rising, while the proportion of the poor NVP remains very high in Wuhan. This study suggested the need to efficiently improve the residential area layout for better urban ventilation to combat COVID-19 without losing other benefits**.** The machine-learning approach can benefit future quantitative research on sick buildings, health surveillance, and housing design.

Study by Leung et al (2024g) "Deep Learning Approach to Forcasting COVID-19 Cases in Residential Buildings' of Hong Kong: A Multi-sources Data Analysis of the Differential Roles of Environments and Sociodemographic during the Emergence and Resurgence of the Pandemic. *ArVix* https://arvix.org/abs/2403.15759"

The current study has demonstrated that time-invariant environmental and sociodemographic factors could add significant value to the rolling historical case counts in forecasting subsequent ones over a two-week horizon across different periods and geographical areas. The impact of environmental and sociodemographic factors on the evolution of COVID-19 highlighted the need to put urban planning and public housing policies at the heart of public health and social policies for curtailing the re-emergence of COVID-19 or the emergence of the next pandemic.

Study by Leung et al 2024h"Analysing the Variations in Emergency Department Boarding and Testing the Transferability of Forecasting Models across COVID-19 pandemic Waves in Hong Kong: Hybrid CNN-LSTM approach to quantitfying building-level sociological risk. *ArXiv*: 2403.13842(cs)"

Study among Medicaid sample in US has shown strong relationship between social determinants of health and ED utlisation (Melissa et al, 2021). Adverse social determinants of health demonstrated strong positive associations with Veteran Health Administration of ED utilization even after accounting for several demographic and health-related factors (Camille et al, 2020). A systematic review has shown that the major determinants of ED utilisation by older adults were largely individual factors in Singapore (Tang et al, 2022). Among the significant individual determinants were predisposing (staying in public rental housing, religiosity, loneliness, poorer coping), enabling (caregiver distress from behavioural and psychological symptoms of dementia) and health factors (multimorbidity in patients with dementia, frailty, primary care visit in last 6 months, better treatment adherence). Evaluation of societal determinants of ED utilisation was lacking in the included studies so this study would have a more holistic examination of the determinants of ED utilisation.

Other factors affecting long waiting time at ED can be related to personal and clinic factors and one study has shown that age, normal mental status and internal medicine admission were independent predictors of prolonged stay (>6 hours) (Siamisang et al, 2020). It is important and worthwhile to investigate further using more advanced models to unfold the social determinants of health so interventions outside healthcare sector would work in synergy with clinical interventions to minimise ED utilisation hence long waiting hours.

Study by Cereletti et al (2021) has shown that the perceived built environment of satisfaction with the apartment and neighbourhood) was associated most strongly with higher scores in both measured health related quality of life (HRQoL) domains after adjusting for variables describing the social environment. Poorer HRQoL will inevitably lead to increase healthcare utilisation such as ED. Study has shown that poor physical HRQoL was associated with increased hazard of time to first ED visit independent of diseases/illnesses so improvement of physical HRQoL particularly elderly, would have the potential to reduce the first ED visit (Nasser et al, 2018). The findings from the current study have further added evidence the importance of perceived built environment and HRQoL in addressing high utilisagtion of healthcare facilities such as ED.

A study by Wong et al in Hong Kong (2018) have found that potential health-related behaviors mediators including smoking status, alcohol drinking, physical activity levels, walking activity and total physical activity were significantly correlated with the overall satisfaction with the neighborhood environment. The health damaged behaviours would have impact of decline of health status leading to high healthcare utilisation such as ED.

The transferability of learning models built utilized in this study between waves four and five demonstrated a recognizable surge pattern throughout, with the period between waves four and five being particularly significant. The study highlights the importance of the built environment and socio-demographic factors in surge capacity, providing valuable insights to enhance healthcare system preparedness for future outbreaks.

Key summary of policy recommendations

(1) The "Young-old" group (above 50 and below 65) should be targeted group for primary, secondary and tertiary prevention as they are more likely to have re-hospitalisation after discharge and more likely to have more intensive interventions in hospitals and greater number of chronic illnesses.

At community level, health promotion programme should be targeted at "Young-old" Group to address how they can avoid risk factors and enhance their daily living conducive to health to protect them from developing chronic illnesses, primary prevention. Many districts have developed "Healthy City initiatives" at district level and should be further consolidated. The District Councils would take the lead to engage key stakeholders in their respective districts to promulgate "Healthy Cities" initiatives to enhance the health of local population particularly the "Young-old" as more attention is usually drawn to older age group.

The ”Young-old” group should be encouraged to become members of District Health Centres/ District Health Centres Express (DHC/DHCE) so they can be screening for any risk factors and/or early stage of chronic illnesses, secondary prevention. They would managed at early stage with intensive lifestyle modification ± medical intervention by local doctors/general practitioners (GPs) as DHC/DHCE has formed a network of GPs in their catchment area. This will enable early treatment to minimise hospitalisation not only for further development of chronic illnesses recently detected and also re-hospitalisation of other co-existing illnesses.

Those “Young-old” with established chronic illnesses should be better supported in community to avoid further complications of the existing illnesses. The DHC/DHCE at locality would serve as local hub to assess and mobilise the local resources to empower those with chronic illnesses to maintain healthy living such as access to healthy diet, maintenance of mobility, compliance with health advice, psycho-social support etc, delaying complications of their existing illnesses.

The physical and mental health needs of the “Young-Old” and also the working class population should not be under-estimated. In order to facilitate DHC to connect them to related supportive services to promote holistic health and well-being, health conditions affecting their active daily living such as pain, depression, insomnia, anxiety should be included as target conditions to be served by DHCs

(2) Notwithstanding SOPC follow up would minmise re-hospitalisation after discharge, this would not be feasible within short period of time. Different types of primary health care service including would be alternative such as GP services, transitional nursing care, community-based rehabilitative programmes

The DHCs and local general practitioners (GPs) should work in synergy as integrated primary healthcare team to target the ambulatory care-sensitive conditions and provide different types of services meeting the needs of the patients at different points of clinical pathway in well co-ordinated fashion rather than in silos. This will allow those discharged cases to be better managed in the community setting.

(3) The CDCC scheme should extend to management of common health conditions discharged from hospitals particularly those aged below 65 as they tend to have more interventions and co-existing chronic illnesses resulting higher rate of re-admission.

The Primary Care Commission would work closely with scholars in Family Medicine/Primary Care and academic Colleges/Institutions to identify the risk of admission for common health conditions encountered in community and map out the management pathways to avoid unnecessary hospital admission. The GP morbidity survey (latest one will be published soon) and the hospital discharge statistics would help to map out those conditions at risk of admission.

(4) For COPD patients, medical service utilisation and comorbidities have been found to associate with survival with greater significance than their lung functions and clinical acuity. COPD would be one example why co-care is needed to minimise in-patient admissions and other medical service utilisation as well as optimal management of co-existing illnesses.

COPD patients should not be regarded as patients with one condition only. Their other existing heath problems also need to be managed. Therefore, they should be included in co-care system for optimal management of their co-morbidities together with their COPD conditions.

(5) COPD illustrates the importance of case management approach to personalized multidisciplinary care that can best facilitate a trans-disciplinary approach to healthcare across the patient journey such as COPD.

Community based pulmonary rehabilitation programmes should be in place in community operated by DHC/DHCE in collaboration with NGOs with the experience in the locality. Local volunteers can be trained as care team with basic skills to facilitate mobilisation of patients who would move around. This basic support can help COPD patients and families to encourage mobility avoiding house bound (being physical active is one natural type of pulmonary rehabilitation) and other basic needs at home.

(6) Population with metabolic indicators such as waist-to-hip ratio and abnormal lipid level as well as presence of co-existing chronic illnesses such as cardio-vascular diseases, diabetes mellitus, chronic kidney disease, COPD should also be monitored for development of cancers such as colorectal, liver, pancreatic bladder, gastric, kidney cancers.

Community based health promotion is very much needed to organise screening programmes to identify risk factors at early stage. This should be part of "Healthy City Initiatives" to educate the whole population on the important link between cancer and various chronic illnesses, e.g., diabetes mellitus. The District Council and District Office would earmark resources for screening programme and follow up interventions.

Residents with chronic illnesses should be monitored for possible risks of development of different types of cancer by their local GPs with support from tertiary care.

(7) The Specific configurations of residential buildings' internal and external environment put their residents at higher risk for poor general health, pain, acute hospital admission, and COVID-19 infections. Built environment features were also assigned greater statistical importance to general health compared to features associated with one's socio-demographics, health, and health-related behaviours and service utilisation, and buildings' built environment features' statistical importance to general health outcome were also shown to predicted cumulative COVID-19 case counts.

(8) Findings from the ED study have further added evidence the importance of perceived built environment and HRQoL in addressing high utilisation of healthcare facilities such as ED.

(9) The impact of housing characteristics on mental health would be more direct in communities with relatively poor housing conditions, and the effect might be indirect for communities with relatively good housing conditions.

Detail social history of residents would be obtained in primary healthcare setting such as GPs and DHCs to identify those patients with high living environmental risks to assess their general health particularly pain, and the interference for effective prevention. Incentives system should be built in for GPs and DHCs to take on this preventive role.

Although it is not possible to modify housing conditions within short period of time, one can consider moblising community resources such as "Care Team" (equipping basic skills to facilate mobility of people requiring basic assistance), "Community Living Room", enhanced accessibility to gyms and recreational facilities to mitigate the impact of adverse living conditions, and also mitigate the potentially high acute medical care needs and COVID-19 infection risk predisposed by specific architectural elements and sociodemographic profiles of public housing.

Community pain screening can be targeted to residential areas whose built environment contributed most to pain interference and informing the design of intervention programs to minimise pain interference

among those suffering from chronic pain and showed specific characteristics. Early intervention would empower patients on pain management to minimise the pain and suffering to prevent further deterioration and restore the functional activity as far as possible.

DHCs having the capacity to serve as primary healthcare hub would have dual roles. Apart from health management of the residents, DHCs would mobilise community resources to help residents coping with health and unfoavourable living conditions. Instrumental homecare, on its own or combined with either one or both of the other home care services, has been shown to yield the greatest costs savings compared to other services. DHC with strong team of social workers would help to streamline the home care services according to the needs of the residents.

A local task force is needed with inputs from building experts to plan for improvement of building configuration in long run.

(10) The short-form of CGA developed into a scoring system would be a potential tool for community screening to prevent frequent hospitalization among the medical system's "super-utilisers".

Most of the items are related to psycho-social resources. Additional investments can be made to ensure that the District Offices are well-informed about the residents in their district, by providing access to appropriate data and data visualisation/decision-support tools. The development and implementation of health and social policies for the districts can benefit from District Offices' unique perspectives on their respective constituents' environment, behavioural patterns, and care needs.

Social services and ambulatory and post-acute services can reduce re-hospitalisation as effective means of tertiary prevention, and data-driven optimization programming, the resultant optimal pairing among care needs of the service end-users, service types, and the timing of service provision can maximize service efficiency and medical cost savings for the district's health and care system.

Health is physical, psychological and social well-being not merely absence of diseases. Various findings from this study have highlighted the physical and psycho-social needs are inseparable. It is time to reform the infrastructure to avoid health services and social services operating in silos. Health and social care should be taken as integrated care under one umbrella for effective and efficient use of scarce resources.

Many behaviours are linked to different types of chronic illnesses and also have impact on prognosis. Profiling the primary, secondary and tertiary prevention needs is needed for development effective primary health care system requires the integration of preventive, promotive, curative, rehabilitative, and palliative healthcare services. We recommend investment in the expertise and decision-support tools to enable data-driven optimization of medical and social services at the district level.

The different studies of this project have shown that people with one chronic illness are more likely to suffer concurrently from psychosocial issues, functional dependency and other co-morbid chronic illnesses. Different studies have also revealed the frequent re-admission of chronic illnesses due to lack of support in the community. Existence of risk factors and chronic illnesses put people at higher risk of developing other morbid conditions such as cancer. Living environment and psycho-social demographic factors play significant roles for hospital admissions. Minimisation of hospitalisation and re-hospitalisation would have significant impact on relieving health burden. This requires actions outside the hospital setting.

Social services and ambulatory and post-acute services can reduce re-hospitalisation as effective means of tertiary prevention. Types and modes of services, target population group, how different services would create synergistic

effects, are all important factors whether greater care in the community can be achieved. One would no longer provide different services in silos with primary healthcare and hospital care operating separately, and health and social care services are distinct services. The data-driven optimization programming, the resultant optimal pairing among care needs of the service end-users, service types, and the timing of service provision can maximise service efficiency and medical cost savings for the district's health and care system. The local District Office and the local District Council can support the local DHC/DHCE to function as health hub for local population in collaboration with key stakeholders and care providers to maximise the resources to deliver effective and efficient care for local population.

Figure 1 outlines a framework to optimise 3 tiers of prevention in the community for adults and how different parties in community would contribute.

Limitation

There are limitations of different studies of this project. Limitations of the current public medical system-based study include the potential applicability of its findings and the decision tool to HK's own private medical system The public hospital covers over 90% of in-patient services. The coding system might not cover all the health conditions of patients basically the chief complaints.

The study on pain and living condition only included its participants recruited from one NGO only and the sample size was not very larger. However, the random forest-based algorithm can select features by integrating individual decision trees' rank-ordering of features based on each feature's unique and combined contribution to the study outcomes, and optimize the cutoff values defining the response levels of selected features assigning the weights to the corresponding response levels of the selected features. The method can accommodate the current sample size. The study on built environment and COVID-19 has limitation only based on public housing estates.

The study on abbreviating CAGA as screening tool has some limitations. The studied sample of homecare service clients may not be representative of the population of community-dwelling elderly. However, the sample of standard government-subsidized community service clients, judged by licensed health professionals as living independently with low clinical and social needs, is similar to clients encountered in the primary and community care settings for which the CGA was abbreviated. There would be recall bias of hospitalization events two years prior to administering the CGA. The self-report hospitalization events were verified by the hospital discharge summaries issued to the study participants.

CHAPTER 6: KNOWLEDGE TRANSLATION/DISSEMINATION

In addition to the usual channel for disseminating our findings via academic journals, key findings will be presented in seminars for different stakeholders. One of the investigators (Albert Lee "AL") was invited as key speaker by Hong Kong Institute of Planners for seminar on "A Two-Part International Series Planning Healthy" on 27 February 2024 with positive feedback locally and overseas. Another forthcoming presentation is planned in July 2024 for the conference of College of Professional and Continuing Education of Hong Kong Polytechnic University focusing on impact of living environment and COVID-19. In forthcoming textbook of primary healthcare co-edited by AL will include the findings of various studies of this project in the chapters of "Philosophy of Primary Healthcare" and "Healthy Cities".

The team has presented the key findings to key stakeholder of Healthy City movement in Hong Kong such as Kwan Tsing Safe and Healthy City Association (KTSHCA) operating K&T DHC, Wong Tai Sin Healthy City Association involved in operation of WTS DHC, Southern District Healthy City Association involved in operation of Southern District DHC, Tusen Wan Healthy City Association closely linked with Tusen Wan DHC, and Island Healthy City Association. The Chairman of KTSHCA is convenor of the group and he is planning press conference to be held to disseminate our findings to officials, professional leaders and community leaders in K&T to kick-start a district-wide initiative on community rehabilitation for COPD and chronic pain. The whole group is interested in making use of the findings to further develop health city movement in the territory.

The team has the privilege of presenting the findings to the District Officer of SSP and his colleagues on SSP's built environment and supporting his partnership with the housing authority, the hospital authority, and the District Health Centre to transform the socioecology of SSP. Consequential to our support to the District Officer, one of the investigators (Eman Leung "EL") was awarded 0.70M to build a predictive AI platform based on the different algorithms that SPPR has funded us to publish for the purpose of 1) informing the District Officer's policy advocacy and decisions, 2) potential adoption at SSP District Health Centre via the District Officer's brokering, and 3) optimizing the deployment of the Care Team at SSP and narrowing down the scope of what Care Team needs to screen and refer that will yield the most transformative primary care outcome. Not only does this collaboration with, and receiving support from the District Officer fulfilled **Objective #4** of the current project, but the platform has also been used for targeted primary and secondary prevention promotion for SSP and K&T residents in events hosted by the District Office of SSP and K&T respectively. Currently, Chairman of KTSHCA is working with the District Health Centre and the K&T District Office to adapt the AI platform for K&T-specific prevention profiles.

Zoom meeting was held with Cluster Chief Executive, Chief of Service and Department Operation Manage of Medicine and Geriatrics, Nursing Consultant for community care of Kowloon West on 1 February 2024. EL and AL presented the key findings to them and will send them a report with key features particularly factors associated with re-hospitalisations so the Hospitals would map out strategies to collaborate community partners to minimise re-admission].

The supervised learning-abbreviation of Comprehensive Geriatric Assessment (CGA) can be developed as community screening tool of elderly at risk of hospital admission for different districts after validation of sample of local elderly population. The above-mentioned Healthy City Associations are eager to use the tools to idenitify those elderly at high risk so they can map out the community resources including local DHCs to support the high risk elderly.

Both AL and EL had meeting with senior staff of Primary Healthcare Office (now Primary Healthcare Commission "PHC") overseeing primary health development to present the research objectives and seeked their

inputs. Meeting with the Commissioner/Assistant Commissioner/Director of DHCs of PHC will be arranged to present the key findings with sugggested recommendations.

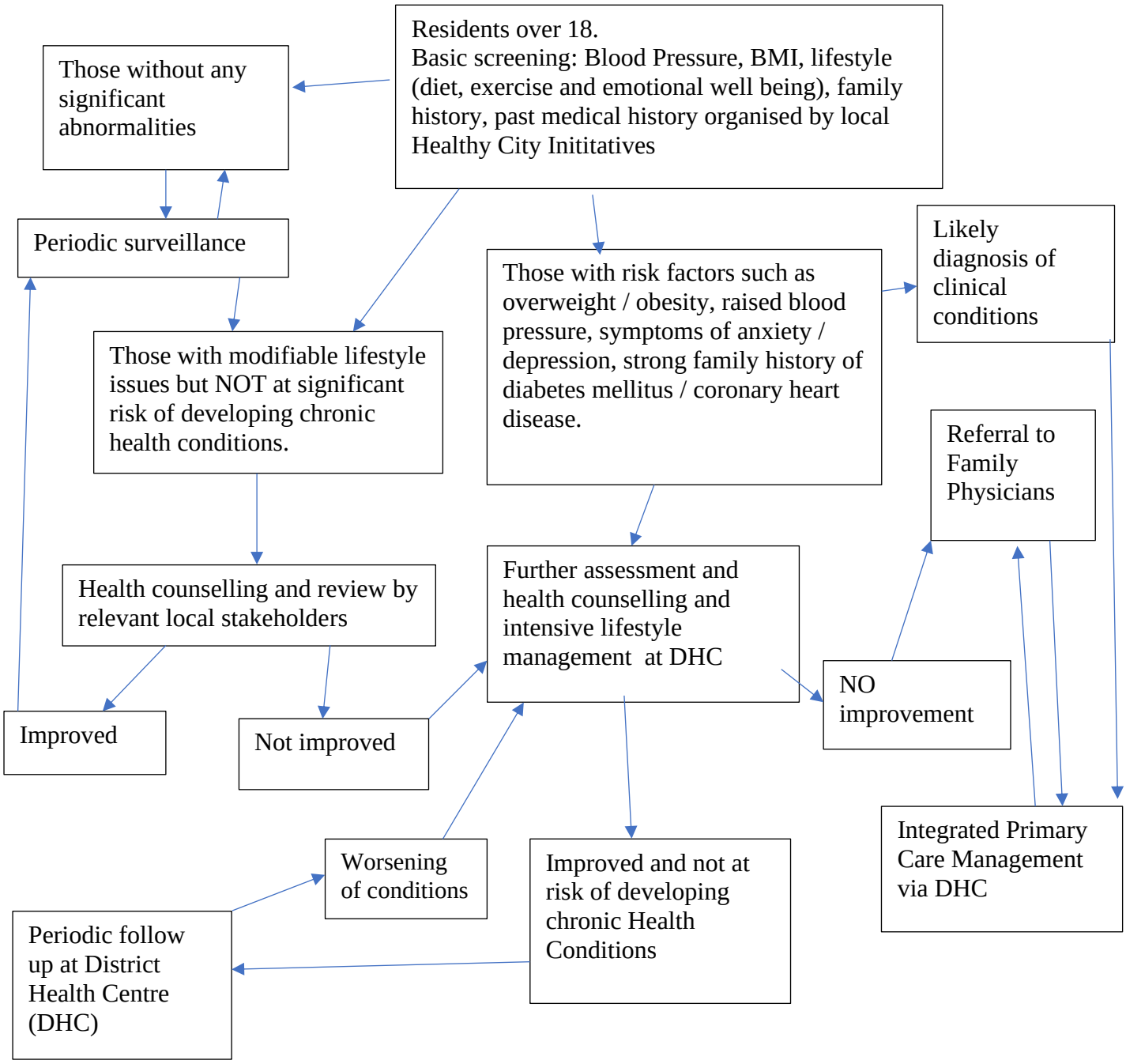
AL will incorporate the relevant contents of this study into teaching materials of "Healthy Setting" of Master of Public Health Programme, "Health Promotion in Practice" of Medical Year 5 community medicine module, "Inter-disciplinary and inter-professional collaboration and communication" of Master of Healthcare Management. AL and EL will make use of the findings to co-edit a book on "Profiling primary, secondary and tertiary prevention needs to improve population health" as part of "Healthy Setting" series (2 books have been published so far with wide audience and Chinese translation is available for one book).

CHAPTER 7: CONCLUSION

Primary prevention is preventing population exposure to risk factors or enhancing their exposure to protective factors. Secondary prevention is early identification of population at risk for early intervention. Tertiary prevention is prevention of deterioration of established health conditions. Prevention guidelines are typically designed for average individuals in the population and developing evidence-based recommendations targeting specific sub-population defined by traditional risk factors is challenging when data on the balance of benefits and harms is lacking. Studies commonly segment their respective sample only with age and specific clinical conditions pre-selected by authors, thus failing to distinguish high-need patients whose age and clinical conditions were the same but differed in the intensity of required care due to, for example, comorbid chronic illnesses; or only with care costs or service utilization parameters totaled across different care locale on a patient journey where they were sourced without distinguishing the variation in clinical acuity and complexity of patients serviced by different care locales. This current project can be showcases of a methodology for, and the benefit of, achieving decision-centric precision public health through the development and deployment of artificial intelligence algorithms across fragmented databases ascertained from multiple systems that were siloed to profile the prevention needs of populations and, accordingly, inform medical-social service optimization.

The artificial intelligence algorithm can map the trajectory of how primary, secondary and tertiary prevention needs to forecast the growing demands of medical and social services of the population. The “one-size-fits-all” approach to multidisciplinary care is not sufficient in addressing the needs of patients with multiple and chronic health conditions, who may encounter many healthcare professionals from different disciplines across primary and secondary care settings. Many behaviours are linked to different types of chronic illnesses and also have impact on prognosis. Profiling the primary, secondary and tertiary prevention needs is needed for develop of an effective primary health care system requiring integration of preventive, promotive, curative, rehabilitative, and palliative healthcare services.

Figure 1. Community Assessment for 3 tiers of Prevention (Adapted from Albert Lee 2023)



References

Abir, M., Hammond, S., Iovan, S. & Lantz, P. M (2019). Why More Evidence Is Needed On The Effectiveness Of Screening For Social Needs Among High-Use Patients In Acute Care Settings. *Heal. Aff. Blog* 1–6.

Al Quait A, Doherty P (2016). Does cardiac rehabilitation favour the young over the old? *Open Heart*, 3(2):e000450

Allard-Ratick M, Khambhati J, Topel M, Sandesara P, Sperling L, Quyyumi A (2018). Elevated HDL-C is associated with adverse cardiovascular outcomes. *Eur Heart J*, 9(1):ehy564.50.

Ahn SV, Lee E, Park B, et al. Cancer development in patients with COPD: a retrospective analysis of the National Health Insurance Service-National Sample Cohort in Korea. *BMC Pulm Med.* 2020;20(1):170.

An C, Choi JW, Lee HS, et al (2021). Prediction of the risk of developing hepatocellular carcinoma in health screening examinees: a Korean cohort study. *BMC Cancer*, 21(1):755.

Behera BK, Prasad R, Shyambhavee. Primary health-care goal and principles. *Healthcare Strategies and Planning for Social Inclusion and Development*. 2022:221-39.

Bertero E, Canepa M, Maack C, Ameri P (2018). Linking heart failure to cancer: background evidence and research perspectives. *Circulation*, 138(7):735-42.

Benjamin Aretz, Gabriele Doblhammer, Fanny Janssen, Effects of changes in living environment on physical health: a prospective German cohort study of non-movers, *European Journal of Public Health* 2019; 29 (6); 1147–1153, https://doi.org/10.1093/eurpub/ckz044

Camille I. Davis, Ann Elizabeth Montgomery, Melissa E. Dichter, Laura D. Taylor, John R. Blosnich. Social determinants and emergency department utilization: Findings from the Veterans Health Administration. *The American Journal of Emergency Medicine* 2020; 38 (9):1904-1909. https://doi.org/10.1016/j.ajem.2020.05.078.

Bushnell, M.C,. Ceko, M., Low, L.A (2013). Cognitive and emotional control of pain and its disruption in chronic pain. *Nat Rev Neurosci*, 14(7):502-11. doi: 10.1038/nrn3516.

Bushnell, M.C., Case, L.K., Ceko, M., Cotton, V.A., Gracely, J.L., Low, L.A., Pitcher, M.H., Villemure, C (2015). Effect of environment on the long-term consequences of chronic pain. *Pain*, 156 Suppl 1(0 1), S42-S49. doi: 10.1097/01.j.pain.0000460347.77341.bd.

Cerletti P, Eze IC, Keidel D, Schaffner E, Stolz D, et al. (2021) Perceived built environment, health-related quality of life and health care utilization. *PLOS ONE* 16(5): e0251251. https://doi.org/10.1371/journal.pone.0251251

Chong JL, Lim KK, Matchar DB (2019). Population segmentation based on healthcare needs: a systematic review. *Syst Rev*, 8(1):202.

Cravo A, Attar D, Freeman D, Holmes S, Ip L, Singh SJ (2022). The Importance of Self-Management in the Context of Personalized Care in COPD. *Int J Chron Obstruct Pulmon Dis*,17:231-43.

Dalal HM, Doherty P, Taylor RS (2015). Cardiac rehabilitation. *BMJ*,351:h5000.

Bourbeau J, Julien M, Maltais F, Rouleau M, Beaupré A, Bégin R, et al (2003). Reduction of Hospital Utilization in Patients With Chronic Obstructive Pulmonary Disease: A Disease-Specific Self-management Intervention. *Archives of Internal Medicine*, 163(5):585-91.

Davis AC, Shen E, Shah NR, Glenn BA, Ponce N, Telesca D, et al (2018). Segmentation of High-Cost Adults in an Integrated Healthcare System Based on Empirical Clustering of Acute and Chronic Conditions. *J Gen Intern Med*, 33(12):2171-9.

Eissens van der Laan MR, van Offenbeek MAG, Broekhuis H, Slaets JPJ. A person-centred segmentation study in elderly care: Towards efficient demand-driven care. *Social Science & Medicine*. 2014;113:68-76

Finkelstein, A., Zhou, A., Taubman, S. & Doyle, J (2020). Health Care Hotspotting — A Randomized, Controlled Trial. *N. Engl. J. Med.* **382**, 152–162.

Grabowski DC, Mor V (2020). Nursing Home Care in Crisis in the Wake of COVID-19. *JAMA*, 324(1):23-4.

Franczyk B, Rysz J, Ławiński J, Rysz-Górzyńska M, Gluba-Brzózka A (2021). Is a high HDL-cholesterol level always beneficial?. *Biomedicines*, 9(9):1083.

Glikson E, Woolley AW (2020). Human trust in artificial intelligence: review of empirical research. *Acad Manag Ann*, 14:627-60.

Guan, Jingjing & Leung, Eman & Kwok, Kin-On & Hung, Chi & Lee, Albert & Chong, Ka & YAM, Carrie & Cheung, Clement & Tieben, Hendrik & Tsang, Hector & Yeoh, EK. (2023). Quantifying The Risk of General Health and Early Covid-19 Spread in Residential Buildings With Deep Learning and Expert-Augmented Machine Learning. *10.2139/ssrn.4432448*

Guan J, Leung E, Kwok KO, Hung CT, Lee A, Chong KC, Yam CHK, Cheing CKM, Hendrik T, Tsang HWH, Yeoh EK (2024). Applying Deep Learning and Expert-augmented Machine Learning to Quantify the Risk of Poor General Health and COVID-19 in Public Housing Residents' Built Environment: The role of socioecology during the first 365 days of COVID-19. Personal communication

Hong SN, Mak IL, Chin WY, Yu EYT, Tse ETY, Chen JY, et al (2023). Age-specific associations between the number of co-morbidities, all-cause mortality and public direct medical costs in patients with type 2 diabetes: A retrospective cohort study. *Diabetes, Obesity and Metabolism*, 25(2):454-67.

Keene D, Price C, Shun-Shin MJ, et al (2014). Effect on cardiovascular risk of high density lipoprotein targeted drug treatments niacin, fibrates, and CETP inhibitors: meta-analysis of randomised controlled trials including 117 411 patients. *BMJ*, 349: g4379.

Kubota N, Fujiwara N, Hoshida Y (2020). Clinical and molecular prediction of hepatocellular carcinoma risk. *J Clin Med*, 9(12):3843.

Khoury MJ, Iademarco MF, Riley WT (2016). Precision Public Health for the Era of Precision Medicine. *Am J Prev Med* [Internet], 50(3):398–401. Available from: https://pubmed.ncbi.nlm.nih.gov/26547538/

Joynt KE, Figueroa JF, Beaulieu N, Wild RC, Orav EJ, Jha AK. Segmenting high-cost Medicare patients into potentially actionable cohorts. *Healthcare*. 2017;5(1):62

Kan, Z.; Kwan, M.-P.; Ng, M.K.; Tieben, H. The Impacts of Housing Characteristics and Built-Environment Features on Mental Health. *International Journal of Environmental Research and Public Health* **2022**, *19*, 5143.

Lafortune L, Béland F, Bergman H, Ankri J. Health status transitions in community-living elderly with complex care needs: a latent class approach. *BMC Geriatr*. 2009;9:6.

Lantz, P. M (2020). "Super-Utilizer" Interventions: What They Reveal About Evaluation Research, Wishful Thinking, and Health Equity. *Milbank Q.* **98**, 31–34.

Lauby-Secretan B, Scoccianti C, Loomis D, et al (2016). Body fatness and cancer--Viewpoint of the IARC Working Group. *N Engl J Med*, 375(8):794-8.

Lee A (2000). Philosophy of Primary Care. In: Fong BYF, Law V and Lee A (eds). *Primary Care Revisited: Interdisciplinary Perspectives for a New Era*. Singapore: Springer Nature.

Lee A. Family Medicine and Community Health Care. In: Fong K and Tong KW (Eds). *Community Care in Hong Kong: Current Practices, Practice-Research Studies, and Future Directions*. Hong Kong: City University Press, 2014.

Lee A (2023). *Sustainable Healthcare system based on Primary Healthcare*. Building a Sustainable Healthcare System. Forum. Organised by Hong Kong Institute Asia Pacific Studies, The Chinese University of Hong Kong (CUHK), Co-organised by Centre for Ageing and Healthcare Management Research, Hong Kong Polytechnic University (HKPolyU), Hong Kong College of Community Health Practitioners, Policy Research@HKIAPA, CUHK, July 2023

Leung E, Lee A, Tsang H, Wong MC (2023). Data-driven service model to profile healthcare needs and optimise the operation of community-based care: a multi-source data analysis using predictive artificial intelligence. *Hong Kong Med J*, 29(6):484-6. Available from: https://doi.org/10.12809/hkmj235154

Leung E, Tsang H, Wong MCS, Ching S, Lam O, He Y, Liu Y, Law CB, Chow YH, Chen FY, JGuan JJ, Lee A (2024a). Driving precision public health research and interventions in Hong Kong with population data: A machine learning-informed joint medical-social perspective. *Hong Kong Medical Journal* 2024 Under revision

Leung E, Guan JJ, Kwok KO, et al (2024b). Residential Buildings' Internal and External Environments and Residents' Sociodemographic Profiles Differentially Predicted Buildings' COVID-19 Cases Accumulated during the Early Phases vs. the Resurgence of the Pandemic: A Deep Learning Approach to Model Development and 3-, 7-, and 14-Day-Horizon Forward- Chaining Validation. Preprint

Leung et al (2024b). Comorbidities and prior medical service utilizations affect the mortality trajectories of Chronic Obstructive Pulmonary Disease (COPD) patients: Stratifying 2- and 5- year survival probability of

113,754 community-dwelling 65+ COPD patients discharged from Hong Kong public hospitals with a random survival forest-based decision tool. To be submitted to *J of Clinical Medicine*

Leung E, Lee A, Liu Y, et al (2024c). Impact of environment on pain among the working poor: Making use of random forest-based stratification tool to study the socioecology of pain interference. *International Journal of Environmental Research and Public Health* 2024 *21*, 179. https://doi.org/10.3390/ijerph21020179

Leung E, Guan J, Zhang Q, Lee A, Tsang H, Ching CC, Yee H, Yeoh EK, Chen FY (2024d). Screening for Frequent Hospitalization Risk among Community-dwelling Elderly between 2016 and 2023: Machine learning-driven item selection, scoring system development, and prospective validation. *Frontier in Public Health* Forthcoming

Leung et al (2024e). The Valuation of Elderly Homecare Services Under a Joint Medical-Social Budget Perspective. *Frontier in Public Health* Forthcoming.

Leung et al (2024f). Data-driven optimization of a population-level allocation of post-discharge services for diverse patient segments to improve patient outcome and maximize medical cost savings: A case for decision-centric precision public health. *Frontier in Public Health* Forthcoming.

Leung et al (2024g) on Article on "Deep Learning Approach to Forcasting COVID-19 Cases in Residential Buildings' of Hong Kong: A Multi-soueces Data Analysis of the Differential Roles of Environments and Sociodemographic during the Emergence and Resurgence of the Pandemic. *ArVix* https://arvix.org/abs/2403.15759

Leung E, Guan JJ, Kwok KO, Hung CT, Chng CC, Tsang H, Yoeh EK, Lee A. (2024h) Article "Analysing the Variations in Emergency Department Boarding and Testing the Transferability of Forecasting Models across COVID-19 pandemic Waves in Hong Kong: Hybrid CNN-LSTM approach to quantitfying building-level sociological risk. *arXiv: 2403.13842(cs)* (http://arvxiv/abs/2403.13842).

Lin, S., Bakker, M. & Leung, E (2022). Robust meal delivery service for the elderly: a case study in Hong Kong. *Sci Rep* **12**, 21546. https://doi.org/10.1038/s41598-022-25924-6

Lin P-J, Zhong Y, Fillit HM, Cohen JT, Neumann PJ (2017). Hospitalizations for ambulatory care sensitive conditions and unplanned readmissions among Medicare beneficiaries with Alzheimer's disease. *Alzheimer's & Dementia*, 13(10):1174-8.

Liu L-F, Tian W-H, Yao H-P (2012). Utilization of health care services by elderly people with National Health Insurance in Taiwan: The heterogeneous health profile approach. *Health Policy*,108(2):246-

Liu, R. Zhou, Y., Liu Y., Guo R., Gao L (2023). Association between living environmental quality and risk of arthritis in middle-aged and older adults: a national study in China. *Frontiers in Public Health*, 11 https://doi.org/10.3389/fpubh.2023.1181625

Mechanic R (2014). Post-acute care--the next frontier for controlling Medicare spending. *N Engl J Med*, 370(8):692-694. doi:10.1056/NEJMp1315607

Lorig K (editor) (2006). *Living a healthy life with chronic conditions : self-management of heart disease, arthritis, diabetes, asthma, bronchitis, emphysema & others*. Bull Publishing Company

Low LL, Yan S, Kwan YH, Tan CS, Thumboo J (2018). Assessing the validity of a data driven segmentation approach: A 4 year longitudinal study of healthcare utilization and mortality. *PLoS One*, 13(4). Available from: https://pubmed.ncbi.nlm.nih.gov/29621280/

Lu H, Xia M, Qin Z, Lu S, Guan R, Yang Y, Miao C, Chen T (2022). The Built Environment Assessment of Residential Areas in Wuhan during the Coronavirus Disease (COVID-19) Outbreak. *Int J Environ Res Public Health*, 25;19(13):7814. doi: 10.3390/ijerph19137814.

Melissa L. McCarthy, Zhaonian Zheng, Marcee E. Wilder, Angelo Elmi, Yixuan Li, Scott L. Zeger (2021), The Influence of Social Determinants of Health on Emergency Departments Visits in a Medicaid Sample. *Annals of Emergency Medicine*, 77 (5): 511-522. https://doi.org/10.1016/j.annemergmed.2020.11.010.

Naseer M, Dahlberg L, Fagerström C (2018). Health related quality of life and emergency department visits in adults of age ≥ 66 years: a prospective cohort study. *Health Qual Life Outcomes*, 16(1):144. doi: 10.1186/s12955-018-0967-y.

Noi E, Rudolph A, Dodge S (2022). Assessing COVID-induced changes in spatiotemporal structure of mobility in the United States in 2020: a multi-source analytical framework. *Int J Geogr Inf Sci*, 36:585-616.

Nnoaham KE, Cann KF (2020). Can cluster analyses of linked healthcare data identify unique population segments in a general practice-registered population? *BMC Public Health*, 20(1):798.
Soares JJF, Sundin Ö, Grossi G (2004). The stress of musculoskeletal pain: A comparison between primary care patients in various ages. *Journal of Psychosomatic Researc*h, 56(3):297-305.
Wolf D, Rhein C, Geschke K, Fellgiebel A (2019). Preventable hospitalizations among older patients with cognitive impairments and dementia. *Int Psychogeriatr*, 31(3):383-91.
Singal AG, Mukherjee A, Elmunzer BJ, et al (2013). Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol*, 108(11):1723-30.
Visseren FLJ, Mach F, Smulders YM, et al. (2021)> ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*, 42(34):3227-37.
Slack A, Yeoman A, Wendon J (2010). Renal dysfunction in chronic liver disease. *Crit Care*, 14(2):214.
Shi C, de Wit S, Učambarlić E, et al (2023). Multifactorial diseases of the heart, kidneys, lungs, and liver and incident cancer: epidemiology and shared mechanisms. *Cancers (Basel)*, 15(3):729.
Siamisang, K. , Tlhakanelo, J. and Mhaladi, B (2020). Emergency Department Waiting Times and Determinants of Prolonged Length of Stay in a Botswana Referral Hospital. *Open Journal of Emergency Medicine*, 8, 59-70. doi: 10.4236/ojem.2020.83007.
Tang XR, Pek PP, Siddiqui FJ, Malhotra R, Kwan YH, Tiah L, Ho AFW, Ong MEH (2022). Determinants of emergency department utilisation by older adults in Singapore: A systematic review. *Ann Acad Med Singap*, 51(3):170-179. doi: 10.47102/annals-acadmedsg.2021437.
Vuik SI, Mayer E, Darzi A (2016). A quantitative evidence base for population health: applying utilization-based cluster analysis to segment a patient population. *Popul Health Metr*, 35(5):769–755.
Wells, N M., Rollings, K. A., Ong Anthony D., Reid M. Carrington (2019). Nearby Nature Buffers the Pain Catastrophizing–Pain Intensity Relation Among Urban Residents With Chronic Pain. *Frontiers in Built Environment* **,** 5 https://doi.org/10.3389/fbuil.2019.00142
Vaida, B (2017). For Super-Utilizers, Integrated Care Offers A New Path. *Health Aff. (Millwood)*. 36, 394–397.
Wong, F.Y., Yang, L., Yuen, J.W.M. *et al* (2018). Assessing quality of life using WHOQOL-BREF: a cross-sectional study on the association between quality of life and neighborhood environmental satisfaction, and the mediating effect of health-related behaviors. *BMC Public Health*, 18, 1113. https://doi.org/10.1186/s12889-018-5942-3
Yau S, Leung E, Wong MCS, et al, (2024a). Scoring system for predicting the risk of liver cancer among diabetes patients: a random survival forest-guided approach). *Liver International*, Under revision after review
Yau S, Leung,E, Wong MCS, et al (2024b) Differential metabolic dysfunction profiles and site-specific risk of obesity-related cancers in patients with diabetes" *BMJ Open* Apr 3;14(4):e082414. doi: 10.1136/bmjopen-2023-082414.