

PoPStat – COVID19: 利用人口金字塔量化对 COVID-19 的易感性

Buddhi Wijenayake^{1*}, Athulya Ratnayake¹, Lelumi Edirisinghe²,
Uditha Wijeratne³, Tharaka Fonseka⁴, Roshan Godaliyadda¹, Samath
Dharmaratne⁵, Parakrama Ekanayake¹, Vijitha Herath¹, Insoha
Alwis^{5,6}, and Supun Manathunga⁷

¹ Department of Electrical and Electronic Engineering, Faculty of Engineering,
University of Peradeniya, Peradeniya, Sri Lanka

² Department of Mathematics, Faculty of Science, University of Ruhuna, Sri Lanka

³ Faculty of Medicine, University of Peradeniya, Sri Lanka

⁴ Multidisciplinary AI Research Center, University of Peradeniya, Peradeniya, Sri
Lanka

⁵ Department of Community Medicine, Faculty of Medicine, University of
Peradeniya, Peradeniya, Sri Lanka

⁶ Australian Centre for Health Services Innovation (AusHSI) and Centre for
Healthcare Transformation, School of Public Health and Social Work, Queensland
University of Technology (QUT), Brisbane, Queensland, Australia

⁷ Division of Experimental Medicine, Department of Medicine, Faculty of Medicine,
McGill University, Canada.

*Corresponding Author, email: e19445@eng.pdn.ac.lk

摘要 了解人口年龄结构如何影响 COVID-19 的负担对于疫情准备至关重要，然而常见的汇总指标如中位年龄忽略了诸如偏度、双峰性和高风险群体的比例权重等关键分布特征。我们扩展了 PoPStat 框架，该框架最初设计用于将整个人口金字塔与特定原因的死亡率联系起来，并将其应用于 COVID-19。使用 2019 年联合国世界人口展望中的年龄—性别分布以及 Our World in Data 记录截至 2023 年 5 月 5 日每百万人口的累计病例和死亡数，我们计算了人群差异度（从一个优化参考金字塔的 Kullback – Leibler 散度）对于超过 180 个国家，并推导出 *PoPStat – COVID19* 作为该散度与对数转换后的发病率或死亡率之间的皮尔逊相关系数。优化选择马耳他的旧偏斜金字塔作为参考，得出案例 ($r = -0.86, p < 0.001, R^2 = 0.74$) 和死亡 ($r = -0.82, p < 0.001, R^2 = 0.67$) 之间存在强负相关。在另外二十个类似的老偏斜参考中进行敏感性测试证实了这些关联对参考选择的稳健性。与人均国内生产总值、基尼系数、人类发展指数、出生时预期

寿命、中位年龄、人口密度、社会人口指数和全民健康覆盖指数等八个标准指标进行基准测试显示, *PoPStat - COVID19* 超过了人均 GDP、中位年龄、人口密度以及几个其他传统指标, 并在致命负担方面优于所有比较对象。因此, *PoPStat - COVID19* 提供了一个简洁且考虑分布的标量, 用于量化人口对 COVID-19 的脆弱性。

Keywords: 人口金字塔, 新冠死亡率, 人口分布, PoPStat, PoPDivergence

1 介绍

COVID-19, 由 SARS-CoV-2 病毒 [10] 引起的具有高度传染性和可能严重的呼吸道疾病, 已导致全球约 710 万例确诊死亡 [30], 尽管其正式紧急阶段已经结束, 但仍持续作为重要的公共卫生威胁。

世界卫生组织于 2020 年 1 月 30 日宣布为国际关注的突发公共卫生事件 (PHEIC), 并于 2020 年 3 月 11 日宣布为大流行病。急性紧急状态在 2023 年 5 月 5 日 [29] 结束, 当时累计死亡人数达到约 690 万例 [30]。

然而, 该病毒继续传播, 截至 2025 年 5 月 25 日, 全球确诊死亡人数已上升至约 710 万例 [30]。

COVID-19 的严重性在各病例和死亡率中表现出显著差异; 然而, 其总体影响强调了它持续存在的全球健康风险。

较高年龄是 COVID-19 疾病严重程度最强的预测因子 [7,21,31,20], 这表现在老年群体中显著更高的病例致死率 (CFR) 和死亡风险比。美国 and 意大利的研究记录了 80 岁以上人群的 CFR 超过 19–27% ≥, 与 20–54 岁年轻人低于 1% 的比率形成鲜明对比 [21,6]。进一步支持这一点的是 [31] 报道的老年群体死亡风险比大于 6。同样地, [14] 报道了老年群体的相对风险比大于 10。

额外来自多个地区的研究表明, 人口统计学在 COVID-19 结果中扮演着重要角色。例如, [5] 发现, 在低和高病死率 (CFR) 的国家之间, 超过一半的病死率差异可以通过年龄结构的不同来解释。在 [9] 中, 许多国家的 COVID-19 死亡病例按年龄段分布的模式紧密反映了随着人们年龄的增长所有原因导致的死亡人数增加。这表明老年人群承担了更重的负担。[22] 发现, 在调整了病例的年龄分布后, 高达 66% 的国家之间 CFR 差异可以得到解释。这导致了一个年龄标准化中位数 CFR 约为 1.9%。最近 (2024 年), [33] 创建了一个覆盖 156 个国家的全球机器学习模型。他们发现, 在“老龄化驱动”的

集群中，主要包括高收入欧洲国家，人口因素如老年人的比例显著影响了更高的病死率。在次国家级分析中，意大利各省进行的 [18] 显示，70 岁及以上人口较多的省份，特别是那些 80 到 90 岁的人口众多的省份，其病例死亡率显著更高。尤其是在医疗系统压力大和环境污染水平高的情况下更是如此。

先前的研究将简单的年龄金字塔类别与 COVID-19 的结果联系起来。例如，将国家分类为“扩张型”（年轻）或“收缩型”（老年），表明老龄化的人口通常有更高的死亡率 [4]，但这种二元方法无法捕捉到年龄结构的逐渐变化。其他研究在多元模型中包括人口金字塔的形状，但仍不能区分人口统计效应与医疗保健或社会经济因素的影响 [2]。在巴基斯坦进行的一项单一国家分析将其较低的死亡率归因于其年轻的人口分布 [12]，但其结论可能不具备普遍性。这些描述性方法缺乏一个统一、可扩展的指标来量化整个年龄金字塔与 COVID-19 负担之间的关系——这是 PoPStat-COVID19 设计填补的一个空白。

中位年龄是分析中常用的一种常见的汇总统计量，如 [32] 中所示，但它无法全面反映人口脆弱性的全貌，因为它未能捕捉到年龄分布的完整结构复杂性和形状。虽然在某些情况下提供了一定的背景价值，但其局限性在于忽略了关键特征，例如偏斜度、双峰性或高风险年龄段在人口金字塔中的比例权重，这可能会忽略与 COVID-19 死亡率相关的精细的人口统计风险概况。

PoPStat，引入于 [8]，通过利用整个人口的年龄-性别分布克服了单值摘要的这些限制。它通过计算每个国家的人口金字塔与优化参考金字塔之间的 Kullback – Leibler 散度与在那些国家观察到的原因特异性死亡率之间的皮尔逊相关性来量化人口脆弱性。在全球分析 [8] 中，涵盖 371 种疾病类别，PoPStat 解释了神经性疾病 ($r = 0.90$) 和肿瘤 ($r = -0.90$) 的跨国死亡率变异超过 80%，超越了传统的社会经济和人口统计指标。

然而，PoPStat 尚未应用于累计的 COVID-19 结果。在本研究中，我们将 PoPStat 适应于 COVID-19 大流行——表示为 PoPStat – COVID19，并解决三个主要目标，

1. **PoPStat-COVID19 的构建** 我们通过选择一个优化的参考金字塔 (Malta)，使其与 COVID-19 发病率和死亡率的相关性最大化，来计算每百万人口的病例和死亡的标量关联，从而推导出 PoPStat – COVID19。
2. **参考选择下的鲁棒性分析。** 我们通过二十个额外的高性能参考金字塔上进行敏感性分析来评估 PoPStat – COVID19 的稳定性，结果显示类似偏斜的年龄结构对于病例和死亡人数产生了类似的关联系数。

3. **与标准指标进行基准比较。**我们将 PoPStat – COVID19 与人均国内生产总值 (GDP)、收入不平等的基尼指数、人类发展指数 (HDI)、出生时预期寿命、人口中位年龄、人口密度、社会-人口统计指数 (SDI) 和全民健康覆盖指数 (UHCI) 进行了比较，表明人口结构信号在解释 COVID-19 负担方面可以与这些传统预测因子相媲美。

2 方法

2.1 预备知识

PoPStat 和 PoPDivergence 这两个指标，如 [8] 所提出的，使多维人口结构转换为标量变量成为可能，从而便于定量分析人口统计数据与特定疾病死亡率之间的关系。PoP-Divergence 是一种基于 Kullback-Leibler (KL) 散度的度量方法，用于量化人口金字塔与优化参考人口结构偏差的程度，如 (1) 所示。

$$\text{PoPDivergence}(i; \omega) = \sum_{a \in A} P_{i,a} \ln \left(\frac{P_{i,a}}{P_{\omega,a}} \right) \quad (1)$$

其中， $P_{i,a}$ 表示国家 i 中年龄组 a 的人口比例，而 $P_{\omega,a}$ 则表示参考国家 ω 中相应的人口比例。指数 $a \in A$ 跨越所有年龄段，并捕获国家 i 的人口金字塔与参考 ω 之间信息论的差异，数值越大表示结构偏差 [8] 更大。

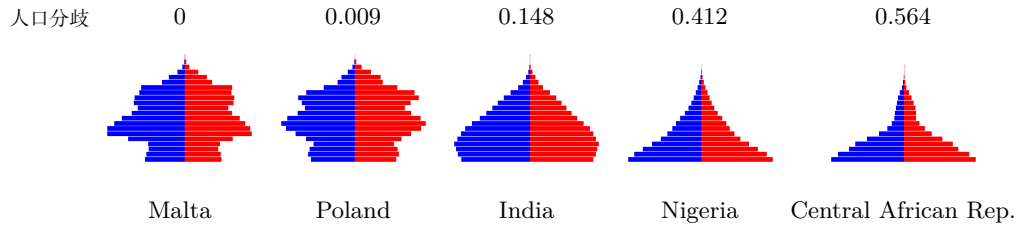


图 1: 偏差的 PoP 发散。马耳他作为参考，表明向进步结构的发展。

此外，人群差异的解释关键取决于参考人群的选择。如图 1 所示，选择马耳他——一个年龄结构相对落后的国家——作为参考，会导致较高的人群差异值对应于相对于马耳他的更年轻（更具扩张性）的人口金字塔。最优的

参考国家是通过参考国调整 [8] 确定的, 该方法选择最大化人群差异和特定原因死亡率自然对数之间的皮尔逊相关性的参考。形式上, 这种选择解决了

$$\omega^* = \arg \max_{\omega \in \Omega} |\text{Cor}[\{\ln S_i\}_{i \in \Omega}, \{\text{PoPDivergence}(i; \omega)\}_{i \in \Omega}]| \quad (2)$$

这里 Ω 表示所有国家的集合, S_i 表示国家 i 的粗死亡率。符号 $\{\text{PoPDivergence}(i; \omega)\}_{i \in \Omega}$ 表示与参考 ω 计算出的一组发散值向量——每个国家一个值——而 $\{\ln S_i\}_{i \in \Omega}$ 则是粗死亡率的对数变换后的向量。然后选择使这两个向量之间的皮尔逊相关性最大的 ω 。这确保了所选参考能够给出人口结构发散与死亡模式之间最强的线性关联。

PoPStat 作为描述人口差异与死亡率之间线性关联的标量总结。一旦通过公式 (2) 选择了最优参考人群 ω^* , 我们就按照 (3) 计算 PoPStat。

$$\text{PoPStat}(\text{ref} : \omega^*) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

其中, x_i 和 y_i 分别表示第 i 个国家的人口结构差异和死亡率的自然对数, 而 $\bar{x}$ 和 $\bar{y}$ 是向量 X (人口结构差异值) 和 Y (对数转换后的死亡率) 的平均值。所得的 PoPStat 值反映了人口结构如何强烈地影响特定疾病的死亡模式。PoPStat 的绝对值越高, 表明人口结构与疾病特异性死亡之间的关联越强。

2.2 数据收集与预处理

我们从联合国 2024 年世界人口展望中获得了 2019 年的 183 个国家层面的人口估算数据, 选择 2019 年是因为它反映了 2019 年底全球出现 SARS-CoV-2 之前的直接人口结构。根据这些数据, 我们构建了按五岁年龄组划分的各国特定人口金字塔, 并将其归一化为性别特定的比例。从 Our World in Data[15] 中获取了每个国家的 COVID-19 病例和死亡人数, 统计截止到 2023 年 5 月 5 日, 这一天世界卫生组织宣布不再将 COVID-19 视为大流行病 [29]。

2.3 构建 PoPStat-COVID19

为了构建 PoPStat – COVID19, 我们将 2019 年人口金字塔与第 2.2 节中概述的累计 COVID-19 监测数据合并, 并应用了方程 2 中的参考国家调整

Algorithm 1: 构建 PoPStat – COVID19

Input: Population proportions $P_{i,a}$ for agegroup $a \in A$ and country $i \in \Omega$, cumulative cases C_i and deaths D_i per million

Output: PoPStat–COVID19_{cases}, PoPStat–COVID19_{deaths}

步骤 1: 计算对数结果;

foreach $i \in \Omega$ **do**

logCases _{i} $\leftarrow \ln(C_i)$;
logDeaths _{i} $\leftarrow \ln(D_i)$;

步骤 2: 参考国家调整;

foreach $\omega \in \Omega$ **do**

foreach $i \in \Omega$ **do**

PoPDivergence($i; \omega$) $\leftarrow \sum_{a \in A} P_{i,a} \ln\left(\frac{P_{i,a}}{P_{\omega,a}}\right)$;

$\rho_{\text{cases}}[\omega] \leftarrow |\text{corr}(\{\text{PoPDivergence}(i; \omega)\}_{i \in \Omega}, \{\log\text{Cases}_i\}_{i \in \Omega})|$;

$\rho_{\text{deaths}}[\omega] \leftarrow |\text{corr}(\{\text{PoPDivergence}(i; \omega)\}_{i \in \Omega}, \{\log\text{Deaths}_i\}_{i \in \Omega})|$;

$\omega_{\text{cases}}^* \leftarrow \arg \max_{\omega \in \Omega} (\rho_{\text{cases}})$;

$\omega_{\text{deaths}}^* \leftarrow \arg \max_{\omega \in \Omega} (\rho_{\text{deaths}})$;

步骤 3: 计算最终的 PoPStat – COVID19;

PoPStat–COVID19_{cases} $\leftarrow \text{corr}(\{\text{PoPDivergence}(i; \omega_{\text{cases}}^*)\}_{i \in \Omega}, \{\log\text{Cases}_i\}_{i \in \Omega})$;

PoPStat–COVID19_{deaths} $\leftarrow \text{corr}(\{\text{PoPDivergence}(i; \omega_{\text{deaths}}^*)\}_{i \in \Omega}, \{\log\text{Deaths}_i\}_{i \in \Omega})$;

程序。每个国家 $\omega \in \Omega$ 都被视为两个对数变换后的严重性度量——每百万人累计病例和每百万人累计死亡——的候选参考金字塔。对于每个候选参考金字塔，我们计算了所有国家的 PoPDivergence，然后计算了 PoPDivergence 与严重性度量之间的皮尔逊相关系数。最大化绝对相关性的金字塔被选为最优参考国家 ω^* 。

由于马耳他被确定为两种情况下的最优参考国家 ω^* ，我们使用马耳他作为参照，通过这些差异与每个结果相关联，并根据公式 (3) 得到了 PoPStat – COVID19。完整的算法概述见算法 1。

2.4 参考金字塔对 PoPStat – COVID19 的影响

为了评估对参考选择的鲁棒性，我们应用了 2 中的调整程序来识别出与每百万人中的对数-COVID-19 死亡人数相关性最极端的十个候选金字塔和相关性最极端的正相关的十个。对于这二十个参考人群中的每一个，PoPStat – COVID19 都重新计算了累计病例和每百万人中的累计死亡人数。

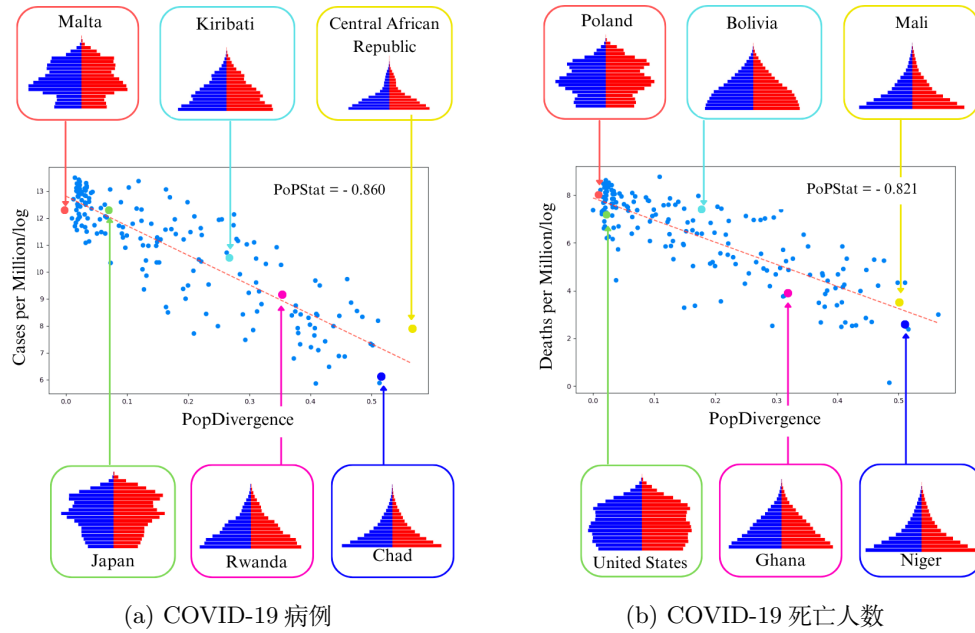


图 2: PoPStat-COVID19 用于马耳他作为参考人口金字塔的 COVID-19 病例和死亡情况

2.5 PoPStat-COVID19 与其它指标的比较

为了将 PoPStat - COVID19 与已建立的发展和人口统计指标进行基准测试,我们编制了 2019 年的国家层面的指标,即人均 GDP [23],基尼系数 [24],人类发展指数 (HDI) [27],出生时预期寿命 [28],中位年龄 [26],人口密度 [25],社会人口指数 (SDI) [11],以及全民健康覆盖指数 (UHCI) [1]。对于除基尼系数以外的所有指标,数据可用于 PoPStat - COVID19 所用的全部 183 个国家 Ω ; 只能获取到 95 个国家的基尼系数值。然后,我们计算了新冠肺炎病例和死亡与每个社会经济指标之间的皮尔逊相关系数,并拟合了简单的线性回归以比较由 PoPStat - COVID19 解释的变化比例 (R^2) 相对于这些传统协变量的情况。

3 结果

3.1 PoPStat – COVID19 与 COVID-19 负担的关联

图 2 说明了 COVID-19 的负担如何随人口结构与马耳他的人口金字塔差异而变化。在左侧面板中, $PoPStat - COVID19_{cases}$ (PoPDivergence 与每百万人病例数之间的皮尔逊相关系数) 是 -0.860 ($p < 0.001$), 表明年龄结构与马耳他差异较大的国家通常报告的病例数量显著较少。在右侧面板中, $PoPStat - COVID19_{deaths}$ (PoPDivergence 与每百万人死亡数之间的皮尔逊相关系数) 是 $r = -0.821$ ($p < 0.001$), 证实了对于死亡率也存在类似的强负相关性。

3.2 对参考人群选择的鲁棒性

为了验证 PoPStat – COVID19 并非仅使用马耳他作为参考的产物, 我们重新计算了统计数据, 采用如 2.4 节讨论的二十个替代参考。与对数死亡率产生最强负的相关性的十个是老龄化严重、呈退化型的人口金字塔, 而与之产生最强正的相关性的十个则是年轻化严重、呈进步型的人口金字塔。表 1 总结了两种结果的系数。

3.3 PoPStat – COVID19 与传统指标的基准比较

为了将 PoPStat – COVID19 与传统的社会经济和人口统计指标进行基准测试, 我们计算了每个指标与每百万的对数转换后的 COVID-19 病例和死亡人数之间的皮尔逊相关系数。表 2 汇总了每个指标的皮尔逊相关系数、双边 p 值、95% 置信区间以及解释的方差 (R^2)。完整的散点图和普通最小二乘拟合如图 3 所示用于病例, 图 4 用于死亡人数, 提供了每个关系斜率陡峭程度、离散度和异常值的可视化比较。

4 讨论

COVID-19 大流行揭示了各国在健康影响和结果方面存在显著差异, 其中许多差异可以归因于多种社会经济因素的差异。一个国家的人口金字塔是一种描述性的人口统计指标, 它以可视化的方式代表了一个国家的年龄分布。它体现了各种经济社会决定因素的累积表现, 反映了卫生保健系统、社

表 1: PoPStat – COVID19 在参考金字塔中的稳健性

参考文献	案例			死亡		
	PoPStat	<i>p</i>	95% CI	PoPStat	<i>p</i>	95% CI
回归 (旧偏斜) 引用						
Malta	-0.860	< 0.001	(-0.89, -0.82)	-0.821	< 0.001	(-0.86, -0.77)
Poland	-0.852	< 0.001	(-0.89, -0.81)	-0.819	< 0.001	(-0.86, -0.76)
Ukraine	-0.853	< 0.001	(-0.89, -0.81)	-0.818	< 0.001	(-0.86, -0.76)
Hungary	-0.849	< 0.001	(-0.89, -0.80)	-0.818	< 0.001	(-0.86, -0.76)
Slovakia	-0.849	< 0.001	(-0.89, -0.80)	-0.818	< 0.001	(-0.86, -0.76)
Switzerland	-0.850	< 0.001	(-0.89, -0.80)	-0.817	< 0.001	(-0.86, -0.76)
Austria	-0.850	< 0.001	(-0.89, -0.80)	-0.817	< 0.001	(-0.86, -0.76)
Estonia	-0.847	< 0.001	(-0.88, -0.80)	-0.816	< 0.001	(-0.86, -0.76)
Serbia	-0.846	< 0.001	(-0.88, -0.80)	-0.816	< 0.001	(-0.86, -0.76)
进步 (偏向年轻) 的参考文献						
Central African Republic	0.775	< 0.001	(0.71, 0.83)	0.701	< 0.001	(0.62, 0.77)
Uganda	0.764	< 0.001	(0.69, 0.82)	0.697	< 0.001	(0.61, 0.77)
Chad	0.764	< 0.001	(0.69, 0.82)	0.696	< 0.001	(0.61, 0.77)
Niger	0.767	< 0.001	(0.70, 0.82)	0.694	< 0.001	(0.61, 0.76)
Mali	0.764	< 0.001	(0.69, 0.82)	0.694	< 0.001	(0.61, 0.76)
Somalia	0.764	< 0.001	(0.69, 0.82)	0.692	< 0.001	(0.61, 0.76)
Burundi	0.757	< 0.001	(0.69, 0.81)	0.690	< 0.001	(0.60, 0.76)
Afghanistan	0.756	< 0.001	(0.69, 0.81)	0.687	< 0.001	(0.60, 0.76)
Mozambique	0.757	< 0.001	(0.69, 0.81)	0.686	< 0.001	(0.60, 0.76)
Malawi	0.751	< 0.001	(0.68, 0.81)	0.684	< 0.001	(0.60, 0.76)

会政策、生育和死亡模式、教育水平以及人口转变模式。因此，可以合理地认为人口金字塔可能与 COVID-19 大流行的影响相关。现有文献证实了人口金字塔的形状对每个群体如何应对 COVID-19 大流行具有强烈影响。^[4]

本研究提出了一种机制，用于推导一种统计指标 PoPStat-COVID-19，该指标通过利用国家人口金字塔中编码的人口结构，显示出与 COVID-19 严重程度的高度相关性。该指标是通过对一个国家的人口分布和优化的参考金字塔之间的 Kullback-Leibler 散度进行比较而得出的。因此，它将金字塔的形状量化为一个标量变量，封装了一个国家的人口特征与其最优结构的偏离程度。

我们的研究表明, PoPDivergence 与 COVID-19 发病率 ($r=-0.860$, $p<0.001$) 和死亡率 ($r=-0.821$, $p<0.001$) 之间存在强烈且具有统计学意义的负相关关

表 2: 指标对 COVID-19 病例和死亡与 PoPStat-COVID19 相关性的比较及方差解释

Indicator	案例				死亡			
	r	p	95% CI	R^2	r	p	95% CI	R^2
GDP per capita[23]	0.60	<0.001	(0.49, 0.69)	0.36	0.38	<0.001	(0.24, 0.51)	0.15
Gini index[24]	- 0.33	<0.001	(- 0.52, - 0.11)	0.11	- 0.15	0.21	(- 0.37, 0.08)	0.02
HDI[27]	0.90	<0.001	(0.86, 0.92)	0.80	0.78	<0.001	(0.71, 0.83)	0.61
Life expectancy[28]	0.83	<0.001	(0.78, 0.87)	0.69	0.72	<0.001	(0.64, 0.78)	0.51
Median age[26]	0.83	<0.001	(0.78, 0.87)	0.69	0.77	<0.001	(0.70, 0.82)	0.59
Population density[25]	0.14	0.07	(- 0.01, 0.28)	0.02	0.00	0.96	(- 0.14, 0.15)	0.00
SDI[11]	0.86	<0.001	(0.81, 0.89)	0.74	0.76	<0.001	(0.69, 0.82)	0.57
UHCI[1]	0.85	<0.001	(0.80, 0.89)	0.72	0.78	<0.001	(0.72, 0.84)	0.62
PoPStat-COVID19	-0.86	<0.001	(- 0.89, - 0.82)	0.74	-0.82	<0.001	(- 0.86, - 0.77)	0.67

系，其中马耳他作为参考国家。这表明，那些人口结构与典型的以老年人为主的马耳他相差较大的国家经历了显著较低的 COVID-19 发病率和死亡率。

这些结果与先前的研究一致，表明人口金字塔呈倒置趋势的国家，即老年人比例较大的国家，表现出更严重的 COVID-19 结果。[3] 这是意料之中的，因为高龄是 COVID-19 疾病和死亡率最强的独立预测因素。[13] 较老的人群倾向于有更高的共病比率，并且需要更密集的医疗护理。[16] 这会加剧对 healthcare systems 的压力并提高总体死亡风险。

相反，人口金字塔呈进步型的国家，尤其是非洲地区，年轻人占较大比例，显示出不同的动态。尽管这些国家在年轻一代中病毒传播速度较快，但临床负担较低，病情较轻且住院率也较低。[4] 因此，尽管存在病毒快速传播的可能性，这种人口结构使得总体上对卫生系统的负担减少，起到了某种人口缓冲作用。

在我们的稳健性分析中，所有十个回归参考都返回了较大的负 PoPStat - COVID19 值。人口特征相似的国家——马耳他、波兰、美国和日本——说明了这种稳定性：尽管它们的年龄-性别分布存在细微差异，但其 PoPStat - COVID19_{cases} 值仅相差约 0.03，PoPStat - COVID19_{deaths} 也相差约 0.03。这种紧密聚类证实了当参考金字塔年老且收缩时，PoPStat - COVID19 对年龄结构的适度变化具有高度稳健性。

相比之下，渐进参考产生了较弱但方向一致的正系数（病例中位数为 $r = 0.46$ ，死亡中位数为 $r = 0.43$ ）。这种不对称性源于 KL 散度根据其在参

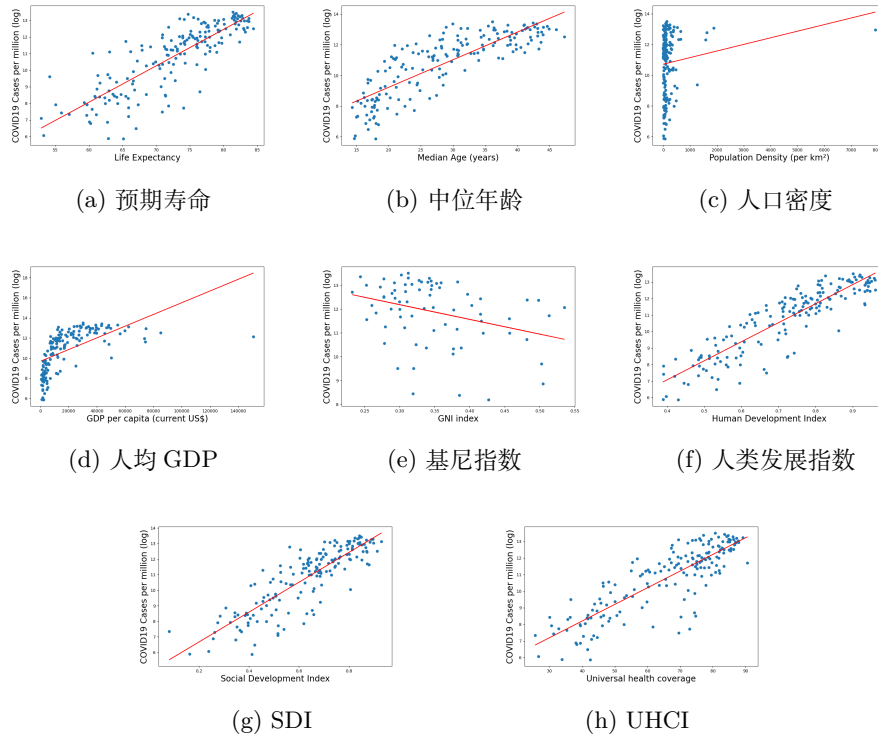


图 3: 每百万人口的对数—新冠肺炎病例与传统指标之间的二元关系

考中的流行程度按比例加权年龄组：一个更老的参考放大了高风险群体的信号，而一个较年轻的参考则削弱了它。重要的是，PoPStat – COVID19 对所有退行性参考以及十个渐进性参考中的八个保持统计显著 ($p < 0.001$)，强调了人口统计学关联的稳健性。

传统的流行病建模指标，如中位年龄、人类发展指数 (HDI) 和人均 GDP，过于简化人口统计特征且存在显著限制。中位年龄忽略了关键的分布特征，例如易感老年人口的比例，导致具有相同中位年龄的人群内部风险不同 [19]。尽管 HDI 与 COVID-19 病例率显示出强烈的关联 ($r = 0.90, R^2 = 0.80$)，但这种关联主要源于广泛的经济社会因素，而不是直接的生物脆弱性。此外，由于报告偏差，特别是在资源不足的情况下，病例率本身极易被低估 [17]。相比之下，PoPDivergence 通过捕捉年龄分布差异直接编码了结构-人口统计学脆弱性，从而为 COVID-19 负担的变化提供了生物学意义的解释。因此，基于 PoPDivergence 并利用完整的疫情前年龄段数据的 PoPStat

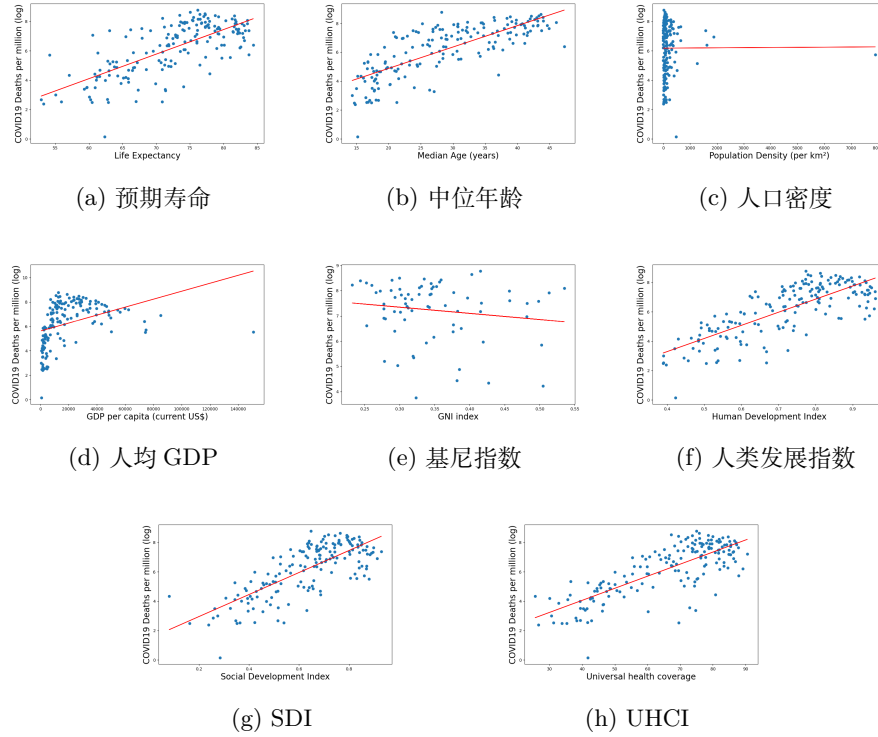


图 4: 新冠肺炎每百万人死亡率与传统指标之间的二元关系

– COVID19 能够可靠地预测病例率 (解释 74% 的方差) 和死亡率 (解释 67% 的方差)。除了 COVID-19 之外, 在存在年龄依赖性风险的情况下, 如流感或呼吸道合胞病毒 (RSV), PoPStat 可以应用于未来的流行病中, 提供了一个用于人口统计学准备评估的现成指标。它的优势在于直接编码人口统计学风险, 这使其不同于基于广泛发展的措施, 例如 HDI, 同时表现相当。

5 结论

我们介绍了一个新的标量度量 $PoPStat - COVID19$, 它捕捉了详细的年龄 – 性别结构对跨国 COVID-19 发病率和死亡率的影响。通过结合完整人口金字塔的 Kullback – Leibler 散度与 Pearson 相关性调整, $PoPStat - COVID19$ 将复杂的 demographic geometry 简化为一个可解释的值。当校准到马耳他的人口金字塔时, 它解释了每百万人口中 74% 的 COVID-19 病例变异和 67% 的死亡率。

在二十个高性能参考金字塔中的稳健性检查表明，当替代参考显示出类似的偏向年长的年龄结构时，*PoPStat - COVID19* 保持其强烈的、统计上显著的相关性。与八个广泛使用的社会经济和人口指标进行比较基准测试显示，由 *PoPStat - COVID19* 隔离的人口结构信号匹配了最佳传统预测器的解释能力，并超过了大多数其他指标。

PoPStat - COVID19 作为人口脆弱性指标提供了几个关键优势。首先，它通过利用人口金字塔的完整年龄分布来提供人口细分，捕捉到了像偏斜和高风险年龄段集中度这样的重要结构性特征，这些特征是中位数年龄等更简单的指标所忽略的。其次，它确保了跨国可比性，在超过 180 个国家内展现出稳健性，无论使用何种特定参考人群，只要该参考人群保持相似的年龄结构即可。第三，它表现出强大的预测能力，解释了 COVID-19 病例和死亡率方差的比例大于广泛使用的指标如人均 GDP、健康覆盖率指数或中位数年龄，因此提供了一种基于生物学并具有统计学强大功能的工具，用于疫情风险评估。

参考文献

1. The universal health coverage coverage index, <https://ourworldindata.org/grapher/universal-health-coverage-index>
2. Birchenall-Jiménez, C.: Centenarians in colombia: a population-based cohort study on the impact of covid 19. *Infectio* pp. 228–234 (Nov 2024). <https://doi.org/10.22354/24223794.1200>, <https://doi.org/10.22354/24223794.1200>
3. Bonanad, C., García-Blas, S., Tarazona-Santabalbina, F., Sanchis, J., Bertomeu-González, V., Fácila, L., Ariza, A., Núñez, J., Cordero, A.: The effect of age on mortality in patients with COVID-19: A meta-analysis with 611,583 subjects. *J. Am. Med. Dir. Assoc.* **21**(7), 915–918 (jul 2020)
4. Cabo, J.M.M.: Demographic variables associated with covid-19 mortality. *Journal of Public Health Research* **9**(4) (Oct 2020). <https://doi.org/10.4081/jphr.2020.1827>, <https://doi.org/10.4081/jphr.2020.1827>
5. Dudel, C., Riffe, T.: Monitoring trends and differences in covid-19 case-fatality rates using decomposition methods: Contributions of age structure and age-specific fatality. *PLOS ONE* **15**(9), e0238904 (2020). <https://doi.org/10.1371/journal.pone.0238904>, <https://doi.org/10.1371/journal.pone.0238904>
6. Edward Livingston, K.B.: Coronavirus disease 2019 (covid-19) in italy. *JAMA* **323**(14), 1335 (Apr 2020). <https://doi.org/10.1001/jama.2020.4344>, <https://doi.org/10.1001/jama.2020.4344>

7. Fei Zhou, T.Y.: Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The Lancet* **395**(10229), 1054–1062 (Mar 2020). [https://doi.org/10.1016/s0140-6736\(20\)30566-3](https://doi.org/10.1016/s0140-6736(20)30566-3), [https://doi.org/10.1016/s0140-6736\(20\)30566-3](https://doi.org/10.1016/s0140-6736(20)30566-3)
8. Fonseka, T., Wijenayake, B.: Devising popstat: A metric bridging population pyramids with global disease mortality (2025), <https://arxiv.org/abs/2501.11514>
9. Goldstein, J.R., Lee, R.D.: Demographic perspectives on the mortality of covid-19 and other epidemics. *Proceedings of the National Academy of Sciences of the United States of America* **117**(36), 22035–22041 (2020). <https://doi.org/10.1073/pnas.2006392117>, <https://doi.org/10.1073/pnas.2006392117>
10. Huang, C.: Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet* **395**(10223), 497–506 (Feb 2020). [https://doi.org/10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5), [https://doi.org/10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5)
11. Institute for Health Metrics and Evaluation (IHME): Global health data exchange (2024), <https://ghdx.healthdata.org/search/site/SDI>
12. Khan, F.: A wide-base population pyramid and other factors explain low mortality in pakistan during covid-19 pandemic. *Journal of Ayub Medical College Abbottabad* **34**(3), 581–584 (Jun 2022). <https://doi.org/10.55519/jamc-03-9444>, <https://doi.org/10.55519/jamc-03-9444>
13. Kofahi, H.M., Swedan, S.F., Khabour, O.F., Nimer, R.M.: Predictors of covid-19 severity and hospitalization: A survey-based study from jordan. *Informatics in Medicine Unlocked* **31**, 100994 (2022). <https://doi.org/https://doi.org/10.1016/j.imu.2022.100994>, <https://www.sciencedirect.com/science/article/pii/S235291482200137X>
14. Lin Fu, J.F.: Influence factors of death risk among covid-19 patients in wuhan, china: a hospital-based case-cohort study (2020). <https://doi.org/10.1101/2020.03.13.20035329>, <https://doi.org/10.1101/2020.03.13.20035329>
15. Mathieu, E., Ritchie, H., Rod s-Guirao, L., Appel, C., Gavrilov, D., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., Roser, M.: Covid-19 pandemic. *Our World in Data* (2020), <https://ourworldindata.org/coronavirus>
16. Mueller, A.L., McNamara, M.S., Sinclair, D.A.: Why does COVID-19 disproportionately affect older people? *Aging (Albany NY)* **12**(10), 9959–9981 (may 2020)
17. Pei, S.: Burden and characteristics of covid-19 in the united states during 2020. *Nature* **598**(7880), 338–341 (Oct 2021). <https://doi.org/10.1038/s41586-021-03914-4>, <https://doi.org/10.1038/s41586-021-03914-4>

18. Perone, G.: The determinants of covid-19 case fatality rate (cfr) in the italian regions and provinces: An analysis of environmental, demographic, and health-care factors. *Science of The Total Environment* **755**, 142523 (Feb 2021). <https://doi.org/10.1016/j.scitotenv.2020.142523>, <http://dx.doi.org/10.1016/j.scitotenv.2020.142523>
19. Prem, K.: The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health* **5**(5), e261–e270 (May 2020). [https://doi.org/10.1016/s2468-2667\(20\)30073-6](https://doi.org/10.1016/s2468-2667(20)30073-6), [https://doi.org/10.1016/s2468-2667\(20\)30073-6](https://doi.org/10.1016/s2468-2667(20)30073-6)
20. Ruchong Chen, W.L.: Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in china. *Chest* **158**(1), 97–105 (Jul 2020). <https://doi.org/10.1016/j.chest.2020.04.010>, <https://doi.org/10.1016/j.chest.2020.04.010>
21. Stephanie Bialek, E.B.: Severe outcomes among patients with coronavirus disease 2019 (covid-19) — united states, february 12 – march 16, 2020. *MMWR. Morbidity and Mortality Weekly Report* **69**(12), 343–346 (Mar 2020). <https://doi.org/10.15585/mmwr.mm6912e2>, <https://doi.org/10.15585/mmwr.mm6912e2>
22. Sudharsanan, N., Didzun, O., Bärnighausen, T., Vollmer, S.: The contribution of the age distribution of cases to covid-19 case fatality across countries: A nine-country demographic study. *Annals of Internal Medicine* **173**, 714–720 (2020). <https://doi.org/10.7326/M20-2973>, <https://doi.org/10.7326/M20-2973>, [Epub 22 July 2020]
23. The World Bank: Gdp per capita (constant 2015 us\$) (2024), <https://data.worldbank.org/indicator/NY.GDP.PCAP.KD>
24. The World Bank: Gini index (world bank estimate) (2024), <https://data.worldbank.org/indicator/SI.POV.GINI>
25. The World Bank: Population density (people per sq. km of land area) (2024), <https://data.worldbank.org/indicator/EN.POP.DNST>
26. United Nations, Department of Economic and Social Affairs, Population Division: World population prospects 2024: Standard projections (2024), <https://population.un.org/wpp/downloads>, most used datasets. Accessed: [Insert Date]
27. United Nations Development Programme: Human development report data center (2024), <https://hdr.undp.org/data-center>
28. World Health Organization: Life expectancy at birth (years) (2024), [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/life-expectancy-at-birth-\(years\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/life-expectancy-at-birth-(years))

29. World Health Organization: News – world health organization. <https://www.who.int/news> (2025), accessed: 2025-06-22
30. World Health Organization: Who covid-19 dashboard (2025), <https://covid19.who.int>
31. Wu, C.: Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, china. *JAMA Internal Medicine* **180**(7), 934 (Jul 2020). <https://doi.org/10.1001/jamainternmed.2020.0994>, <https://doi.org/10.1001/jamainternmed.2020.0994>
32. Xue-Qiang Wang, G.S.: Association between ageing population, median age, life expectancy and mortality in coronavirus disease (covid-19). *Aging* **12**(24), 24570–24578 (Dec 2020). <https://doi.org/10.18632/aging.104193>, <https://doi.org/10.18632/aging.104193>
33. Zhou, C., Åsa M. Wheelock, Zhang, C., Sundin, A., Chen, Z., Jørgensen, N.P., Song, J., Glaab, E., Jørgensen, P.S.: Country-specific determinants for covid-19 case fatality rate and response strategies from a global perspective: an interpretable machine learning framework. *Population Health Metrics* **22**(10) (2024). <https://doi.org/10.1186/s12963-024-00330-4>, <https://doi.org/10.1186/s12963-024-00330-4>