Enhanced Charlson Comorbidity Index varies more among older people in four U.S. health systems

(Title has 94 characters including spaces; same as Short title.)

Joel E. Cohen[1*], T.J. Lin[2], Alice Eggleston[3], Mirta Milanes[3], Amro Hassan[3], Andrea Cassells[2], Jonathan N. Tobin[2,4]

[1] Laboratory of Populations, The Rockefeller University, New York, NY, United States of America; Additional affiliations: Department of Statistics; Department of Earth and Environmental Sciences; Department of International and Public Affairs, Climate School; Columbia University, New York, NY, United States of America; and Department of Statistics, University of Chicago, Chicago, IL, United States of America

[2] Clinical Directors Network (CDN), New York, NY, United States of America

[3] Alliance Chicago, Chicago, IL, United States of America

[4] Co-Director, Community Engaged Research, Center for Clinical and Translational Science, The Rockefeller University, New York, NY, United States of America

* Corresponding author

Email: cohen@rockefeller.edu

Abstract

The enhanced Charlson Comorbidity Index (eCCI) combines a person's concurrent chronic medical conditions (comorbidities) in 39 categories into an index. We show that the average eCCI and the variance of eCCI of people of each age increase with age in de-identified electronic health records of 238,156 adults in low-income communities served by four health systems (2 in Chicago, 2 in NYC). The variance of eCCI approximates a power function of the mean eCCI. This quantitative relationship, not previously recognized, approximates Taylor's power law of fluctuation scaling. The quintiles of eCCI diverge with increasing age, consistent with an increasing variance. Within almost every age group, the frequency distribution of eCCI per individual is well approximated by a negative binomial distribution. This frequency distribution of eCCI per individual of a given age can account approximately for the relation of variance to mean across ages. The increase with age in the mean and the variance of eCCI, and the faster increase of the variance compared to the mean, show that aging brings more comorbidity on average and greater differences in comorbidity among individuals of more advanced ages. Clinical practice should recognize and respond to the greater differences in comorbidity among older people.

Introduction

According to the "World Report on Ageing and Health" (Beard et al. 2016, p. 2146), "ageing is … associated with an increased risk of a person having more than one disorder at the same time (multimorbidity)." Moreover, "the older the age group, the greater the variation found in cognition, physical and sensory function, and social engagement" (Santoni et al. 2015, p. 2).

Hultsch et al. (2002) distinguished three types of variability in the reaction time of younger and older adults: "variability between persons (diversity), variability within persons across tasks (dispersion), and variability within persons across time (inconsistency)." In the present cross-sectional study, each individual is assessed once only by chronological age and an enhanced Charlson Comorbidity Index (eCCI), an indicator of multimorbidity. This structure of the data gives no opportunity to measure inconsistency (which would require longitudinal data) or dispersion (which would require comparing different indices of multimorbidity). We analyze only the mean, variance (Hultsch's "diversity"), quintiles, and frequency histogram of eCCI within groups of people of the same chronological age. The quantitative relationship between the average (or mean) and the variance of eCCI among adults as a function of age appears not to have been examined (e.g., Divo et al. 2014, Chen et al. 2022, Chen et al. 2023, Chowdhury et al. 2023, Bensken et al. 2024).

Before examining data, we predicted that the eCCI might obey a pattern widely observed in ecology and known as Taylor's law (L. R. Taylor 1961, R. A. J. Taylor 2019). Taylor's law predicts that the variance of eCCI should approximate some power of the mean of eCCI across the entire range of adult ages. The data here partly confirm and partly reject this predicted power-law relationship of variance to mean.

Data and definitions

Four health systems (HS 1, HS 2, HS 3, HS 4) in the United States measured the age and the enhanced Charlson Comorbidity Index (eCCI) (Charlson et al. 2025), based on electronic health records, of 238,156 adults in low-income communities. The eCCI used here was designed to anticipate risk of hospitalization and disability whereas several earlier versions of CCI were designed to anticipate risk of death (Charlson et al. 1994, Charlson et al. 2022, Charlson and Wells 2022; see also Ly et al. 2025 and the on-line calculator at https://www.mdcalc.com/calc/3917/charlson-comorbidity-index-cci). The eCCI of a patient was the sum of the points assigned in Table 1 to each chronic condition in the patient's electronic medical record. These 39 conditions generate 2^{39} = 549,755,813,888 or almost 550 billion possible combinations of conditions. The data analyzed here were accessed for research purposes on 2024-12-28.

Table 1. Chronic conditions and assigned points in the enhanced Charlson Comorbidity Index for Tipping Points Conditions and Weights (Charlson et al., 2025).

Chronic Condition	**Comorbidity Weight**
Myocardial infarction	1
Congestive heart failure	1
Peripheral vascular disease or bypass	1
Cerebrovascular disease or transient ischemic disease	1
Hemiplegia	2
Pulmonary disease/asthma	1
Diabetes	1
Diabetes with end organ damage	2
Moderate to severe renal disease	2
Mild liver disease	2
Severe liver disease	3
Gastric or peptic ulcer	1
Cancer	2
Rheumatic or connective tissue disease	1
Hypertension	1
Skin ulcers/cellulitis	2
Depression	1
Warfarin	1
Inflammatory bowel disease	1
Sickle cell disease	3
Hemophilia	3
Muscular dystrophy	2
Cystic fibrosis	3
Tay Sachs	3
Cerebral palsy	2
Uncontrolled seizures	3
Dementia or Alzheimer’s	1
Dialysis	2
Developmental delay	2
Mental retardation	2
Down’s syndrome	3
Bipolar disease	3
Antipsychotics	3
Drug or alcohol addiction	3
Schizophrenia	3
Autism	3
Metastatic solid tumor	6
HIV or AIDS	6
Any transplant: Renal, heart, liver, bone marrow, or lung	6

Age (in whole years plus days from the date of birth to the date on which eCCI was computed) was rounded to the nearest whole year. Single years of age with 100 or more people were selected for further analysis because age groups with few people (here, only the oldest elderly) were judged to give unreliable estimates of the mean and variance (Taylor et al., 1988).

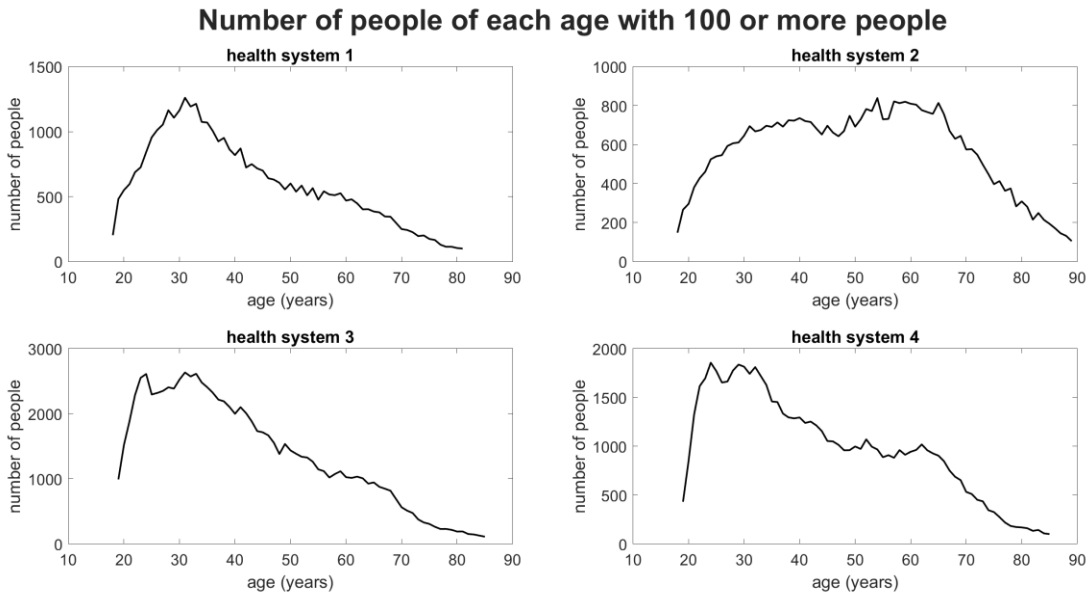
Two of the four health systems (HS 1 = NYC-a and HS 2 = NYC-b) were in New York City and two (HS 3 = Chicago-a and HS 4 = Chicago-b) were in Chicago, Illinois. In New York, the two health systems were Family Health Centers at NYU Langone and Community Healthcare Network. In Chicago, the two health systems were Erie Family Health Centers and Friend Health. In neither city is it publicly known which health system was "a" and which was "b." The health systems and the individuals were fully de-identified for this analysis.

RESULTS

Age structure by HS

The frequency histograms of the number of people by age in each HS (Figure 1) show visible differences among the HSs in the modal age. The modal age of HS 2 is by far the oldest.

Figure 1 Number of people of each age with at least 100 people, by health system. Source: NumberOfPeopleEachAge4HSagemin100_20250131-122802.png



To examine whether the differences in age distribution between and among HSs could have been due to sampling fluctuations from a common single underlying distribution, we used non-parametric and parametric tests.

The non-parametric Kruskal Wallis test (Matlab `kruskalwallis`) revealed statistically significant differences in the ranks of the age of individuals in all four HSs. HS 2 had the oldest age distribution, as Figure 1 suggests. Including only HSs 1, 3, 4, or only the pairs of HSs 1 and 3, or HSs 1 and 4, or HSs 3 and 4, the Kruskal Wallis test again revealed statistically significant differences in the distributions of the age of individuals in different HSs. In all cases, *p* values (testing the null hypothesis of no difference in age distribution) were shown as 0, meaning that they fell below some small positive threshold for rounding to 0.

The parametric test of a difference in mean ages between HSs was analysis of variance (Matlab `anova1`). The mean ages, shown in Table 2, are presumably asymptotically normally distributed by the central limit theorem even if the overall age distribution is not normally

distributed. Once again, the *p* value was shown as 0. We again reject the null hypothesis that the age distributions of the four HSs were sampled from the same underlying distribution.

Table 2 summarizes the distributions of ages and eCCI by HS. The mean eCCI and the variance of eCCI in HS 3 and HS 4 (in Chicago) were notably smaller than the mean and variance of HS 1 and 2 (in New York), and HS 2 had notably higher mean age and mean eCCI than the three other HSs. The Discussion offers possible explanations of this difference between cities.

Table 2. Summary statistics of age and eCCI in four health systems.

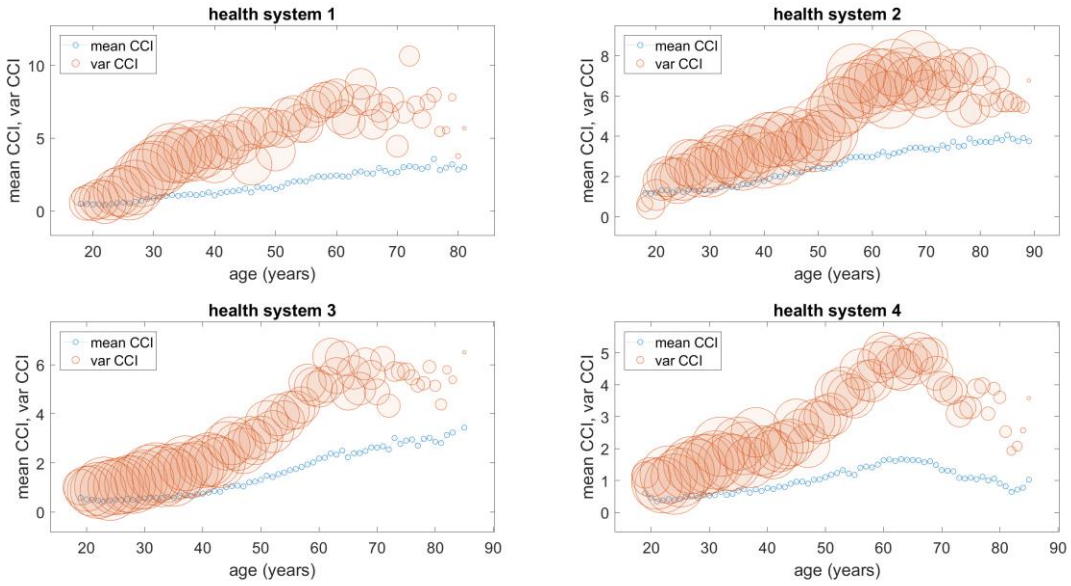
Health system → Attribute ↓	1 NYC-a	2 NYC-b	3 Chicago-a	4 Chicago-b
Number of people	39018	41400	91568	66170
Minimum age years	18	18	19	19
Median age years	39	52	39	40
Maximum age years	106	108	112	110
Mean age	42.3493	51.3300	41.6381	43.4239
Standard dev. age	15.8985	17.7623	15.5392	16.4802
Minimum eCCI	0	0	0	0
Median eCCI	1	2	0	0
Maximum eCCI	24	22	23	21
Mean eCCI	1.4322	2.4736	1.0860	0.8847
Standard dev. eCCI	2.2007	2.3270	1.7543	1.6422

Mean and variance of eCCI by age group

In each HS, for each age with 100 or more individuals, the mean eCCI and the variance of the eCCI of the individuals of that age generally increase with age (Figure 2). For each age, the area (not radius) of the circle for the variance of eCCI is proportional to the number of individuals of that age. The smallest circles at the extreme right of each panel show ages with 100 individuals.

The mean eCCI by age is substantially higher in HS 1 and HS 2 (New York) than in HS 3 and HS 4 (Chicago), and the variance of eCCI is substantially higher in HS 1 and HS 2 than in HS 3 and HS 4. Figure 2 shows clearly that the Poisson distribution (where variance equals mean) could not adequately describe the frequency distribution of eCCI by age group in any HS.

Figure 2 The mean eCCI (blue circles) and the variance (red circles) of the eCCI of the individuals of each single year of age with 100 or more individuals, by HS. The area of each red circle is proportional to the number of individuals of the age corresponding to the abscissa of the center of the circle. Source: MeanCCIvsAgeVarCCIvsAge4HSagemin100_20250131-122802.png



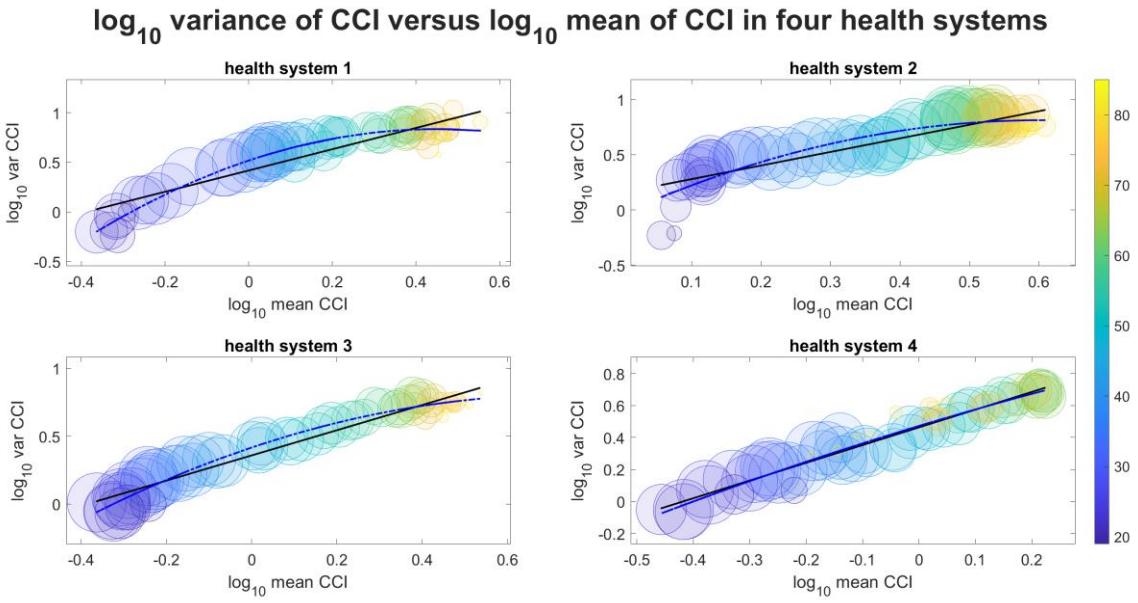
Variance function

The variance function of eCCI in a set of ages shows the variance of eCCI at each age group on the vertical axis as a function of the mean eCCI at each age on the horizontal axis. For each HS, Figure 3 displays the variance function of eCCI by age with at least 100 individuals (circles). The area of each circle is proportional to the number of individuals in the age group. The color of each circle shows the age of the individuals in the group. Figure 3 also displays two models for comparison with the eCCI data: Taylor's law and the quadratic Taylor's law.

Taylor's power law of fluctuation scaling

Taylor's law (TL) was discovered independently at least three times before L. R. Taylor (1961) brought it to wide attention. To describe TL, let $v(a)$ denote the sample variance of eCCI and $m(a)$ the sample mean of eCCI at age a. These quantities are functions of the data, hence the qualification "sample." By contrast with the sample moments, let $\sigma^2(a)$ and $\mu(a)$ denote the population variance and the population mean, respectively, of eCCI in a model of the variance function.

Figure 3 The variance of the eCCI of the individuals of each age with 100 or more individuals as a function of the mean eCCI of each group (center of the circles) and two models: Taylor's law (3) (black solid straight line) and the quadratic Taylor's law (4) (blue dashed curved line). HS 1 and HS 2 are in New York. HS 3 and HS 4 are in Chicago. The area of each circle is proportional to the number of individuals in the corresponding age. The color of each circle represents age from youngest (deep purple or blue) to oldest (yellow). Source: logVarCCIvslogMeanCCITLQTLall4HSagemin100_20250131-140652.png



153

154

155 TL proposes that the variance function is a power law, meaning that the population variance is
156 proportional (with coefficient *c*) to some power (denoted *b* here) of the population mean for a
157 range of values of age *a*:

158 (1) $\sigma^2(a) = c[\mu(a)]^b, c > 0$, for all ages a,

159 and that the sample variance is approximately related to the sample mean by a relationship of the
160 same form with an unspecified error model:

161 (2) $v(a) \approx c[m(a)]^b, c > 0,$ for all ages a.

162 We say that *c* is the coefficient and *b* is the exponent of TL. TL (2) implies a linear relationship
163 on log-log coordinates:

164 (3) $\log_{10} v(a) \approx \log_{10} c + b \log_{10} m(a)$.

165 The so-called quadratic Taylor's Law (Taylor et al. 1978, p. 388, Eq. (14)) (QTL) generalizes TL
166 to a quadratic relationship on log-log coordinates:

(4) $\log_{10} v(a) = c_1 + b_1 \log_{10} m(a) + b_2 [\log_{10} m(a)]^2$,

In each HS, to estimate the values of c, b, c_1, b_1, b_2 from data for ages *a* with at least 100 individuals, we fitted linear (3) and quadratic (4) regressions by least squares using Matlab function `regress` (black solid straight line and blue dashed curve in Figure 3).

Table 3 gives the estimated exponent *b* and estimated coefficient *c* and their 95% confidence intervals (CIs). These confidence intervals are approximate because they assume the mean eCCI is known without sampling error. In HS 1 in New York, the exponent *b* does not differ significantly from 1, while *b* is significantly less than 1 in HS 3 in Chicago and is significantly greater than 1 in HS 3 and HS 4. The coefficient $c = 10^{\log_{10} c}$ significantly exceeds 1 in all four HSs.

Table 3. Least squares estimates of the coefficient *c* and the exponent *b* of sample Taylor's law (3) and nominal 95% confidence intervals based on each age's mean and variance of eCCI. For example, in HS 1, with mean eCCI $m(a)$ and variance $v(a)$ of eCCI in age a, the best approximating TL is $\log v(a) \approx 0.4171 + 1.0729 \cdot \log m(a)$ for all ages a.

health system	exponent_b	low_b	high_b	log10_c	low log10_c	high log10_c
1 NYC-a	1.0729	0.95743	1.1884	0.4171	0.37991	0.45429
2 NYC-b	1.2287	1.0861	1.3713	0.15558	0.096415	0.21475
3 Chicago-a	0.92951	0.87851	0.98052	0.35842	0.3422	0.37464
4 Chicago-b	1.108	1.0442	1.1718	0.46305	0.45062	0.47548

Non-linearity of log variance as a function of log mean

In Figure 3, on log-log coordinates, the variance functions of HS 1, HS 2, and HS 3 are not well approximated as straight lines on log-log coordinates, contrary to TL. Rather, the variance functions of HS 1 and HS 2 appear to be concave.

Table 4 gives the OLS estimates and 95% confidence intervals of the parameters b_2, b_1, c_1 in order from left to right. The estimated exponents b_2 are significantly less than 0 because the 95% confidence interval lies entirely to the left of 0. By contrast, b_2 for HS 4 does not differ significantly from 0.

Table 5 gives the statistics of the TL and QTL regressions.

Table 4. Parameter estimates and 95% confidence intervals for the parameters of the quadratic Taylor's law (4). The substantially and significantly negative estimates of b_2 for New York characterize a concave, non-linear variance function for eCCI.

Health System	b_2	lo b_2	hi b_2	b_1	lo b_1	hi b_1	Intercept c_1	lo c_1	hi c_1
1 NYC-a	-1.5587	-1.8352	-1.2823	1.4061	1.3173	1.495	0.5176	0.48978	0.54542
2 NYC-b	-2.3272	-3.3249	-1.3296	2.8083	2.1197	3.497	-0.03568	-0.13277	0.061417
3 Chicago-a	-0.71316	-0.9095	-0.51683	1.0527	1.0018	1.1036	0.4169	0.39678	0.43702
4 Chicago-b	-0.29897	-0.63715	0.039209	1.061	0.97871	1.1433	0.47221	0.45618	0.48824

Table 5. Statistics of Taylor's law regressions (3) in Table 3 and quadratic Taylor's law regressions (4) in Table 4: coefficient of determination R^2; probability p of the F-statistic of the null hypothesis of no relationship; residual sum of squares RSS. Fcompare is the F-statistic comparing the fit of the quadratic Taylor's law with the fit of Taylor's law; P is the probability of the null hypothesis of no improvement. The null hypothesis is rejected for all four health systems, even though for HS 4 the difference between TL and QTL is visually very small in Figure 3.

Taylor's law regressions (3) in Table 3

Health System	R^2	pFstat2	MSE2	RSS2
1 NYC-a	0.84766	5.09E-27	0.014956	0.9273
2 NYC-b	0.80843	8.02E-27	0.011681	0.81766
3 Chicago-a	0.95394	1.71E-44	0.003822	0.24463
4 Chicago-b	0.94877	1.16E-43	0.002446	0.15898

202

203 quadratic Taylor's law regressions (4) in Table 4

Health System	R^2	pFstat2	MSE2	RSS2	Fcompare	P
1 NYC-a	0.9506	1.44E-40	0.00493	0.30071	127.1	0
2 NYC-b	0.85419	1.41E-29	0.009019	0.62233	21.656	0
3 Chicago-a	0.97492	3.80E-51	0.002115	0.13322	52.688	0
4 Chicago-b	0.95115	1.11E-42	0.002369	0.1516	3.1191	4.72E-06

204

For each HS separately, we tested whether QTL described the data significantly better than TL by an *F*-test:

$$F = \frac{\frac{RSS_{TL} - RSS_{QTL}}{df_{TL} - df_{QTL}}}{\frac{RSS_{QTL}}{df_{QTL}}} \quad (5)$$

where RSS_{TL}, RSS_{QTL} are the residual sums of squares of TL and QTL, respectively, and df_{TL}, df_{QTL} are the corresponding degrees of freedom. Then, if *fcdf* is the cumulative distribution function of the *F* distribution with degrees of freedom df_{TL}, df_{QTL}, we calculate

$$P = 1 - fcdf\left(F, df_{TL}, df_{QTL}\right). \quad (6)$$

If P is larger than a critical value, e.g., 0.05, then QTL is not a significantly better description of the data than TL. If P is smaller than a critical value, then QTL is a significant improvement over TL. The last two columns of Table 5, Fcompare and P, strongly confirm QTL describes the variance functions of all four HSs better than TL, and are significantly concave for HS 1, 2, and 3.

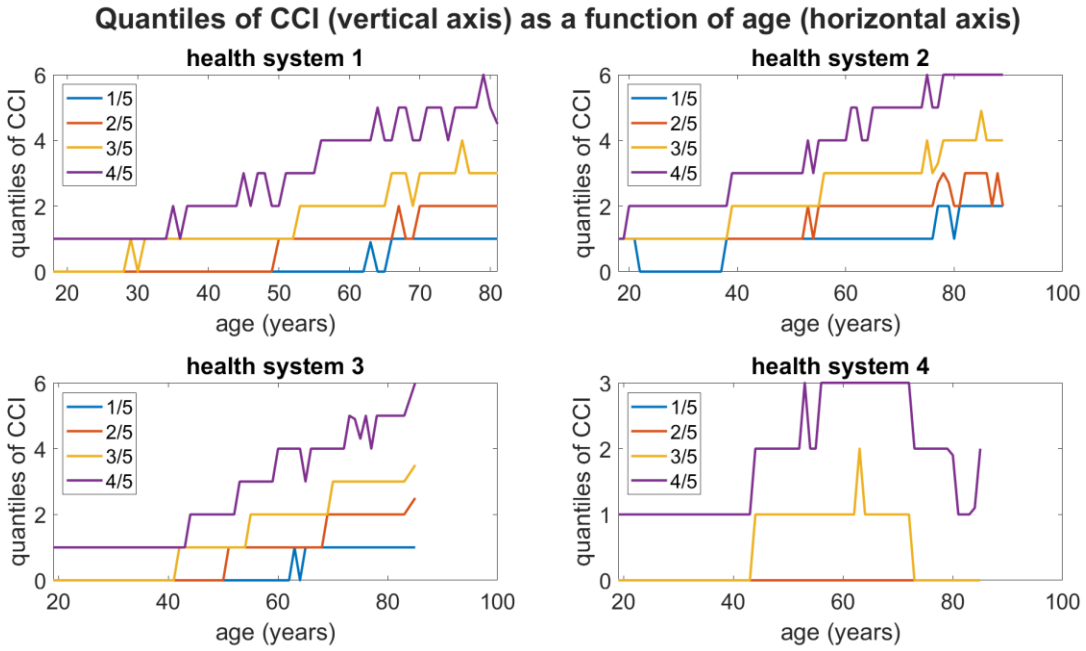
Quantiles of eCCI by age in each health system

The rising trend of the variances of eCCI in Figure 2 indicates increasing scatter ("diversity" in the sense of Hultsch et al. 2002) of eCCI with increasing age. To quantify the increasing scatter further, we display the quantiles of the distribution of eCCI at each age in this section. In the next section, we refine the analysis by fitting the frequency distributions of eCCI at each age to the negative binomial distribution.

The quintiles of eCCI for age *a* are those four values of eCCI such that 1/5, 2/5, 3/5, and 4/5 of the people of age *a* have eCCI less than or equal to those values. These quintiles divide the

ranked list of eCCIs into five equal (or nearly equal) portions. The increases with age in the variance of eCCI suggest that the upper quintiles of the distribution of eCCI should rise more rapidly with age than the lower quintiles because the distribution of eCCI should spread out more with increasing age. Figure 4 confirms this suggestion for each HS and for each age with 100 or more people. The four quintiles of eCCI diverge with increasing age. Figure 4 also shows that the upper quintiles and the entire distribution of eCCI in middle and upper ages are notably lower in HS 4 than in the other three HSs, for reasons yet to be determined.

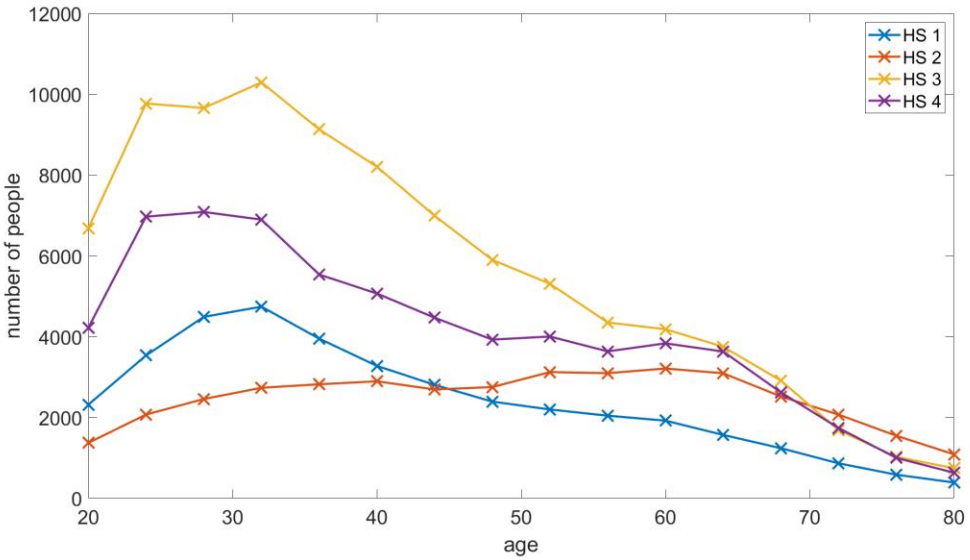
Figure 4 Four quintiles of eCCI as a function of age, for ages with 100 or more people in that health system. Source: QuantilesCCIbyAgeForEachHS_20250128-173406.png



Frequency histogram of eCCI within each age group

The frequency histogram of eCCI within an age group in a HS shows how many people of that age and HS had eCCI equal to 0, how many had eCCI equal to 1, how many had eCCI equal to 2, and so on. To give statistically credible information about the relative frequency of each value of eCCI by age group, we devised 16 age groups containing four years of age each: 19-22 years (that is, from age 20 minus one up to and including 20 plus two), 23-26 (that is, from age 24 minus one up to and including 24 plus two), and so on up to 79-82 (that is, from age 80 minus one up to and including 80 plus two). We label each of these age groups by the age in the group that is divisible by 4, namely, 20, 24, 28, …, 76, 80. Figure 5 shows the number of people in each age group in each HS. As in Figure 1, the larger proportion of people at older ages in HS 2 is visible.

Figure 5 Number of people in each 4-year age group in each health system. Source: NPeopleByAge16GroupsHealthSystem_20250115-184027.png



For all 64 of these age groups (16 age groups for each of 4 HSs), the variance of eCCI exceeded the mean eCCI. (When groups of 100 people or more in individual years of age were used to compare the mean and variance of eCCI, a handful of the youngest and extremely oldest single-year age groups had a variance of eCCI less than the mean eCCI. Evidently, these few exceptions resulted from sampling fluctuations.)

The negative binomial distribution is a classical parametric model of over-dispersed integer-valued nonnegative random variables. Figures 6-9 show the observed counts of people in HS 1-4, respectively, who have each eCCI and the expected counts (green line) from a negative binomial distribution fitted to the empirical probability density function (which is just the frequency histogram divided by the sum of all counts in the histogram). The agreement is generally

excellent. Some deviations arise at the largest values of eCCI where fewer people are observed. Because the vertical axis (number of people) is on a logarithmic scale, the (rare) zero counts are omitted entirely.

Figure 6 For HS 1, the number of people in each 4-year age group (vertical axis of black ×) who have each value of eCCI (horizontal axis of black ×), and the fitted negative binomial distribution (green line). The vertical axis for number of people is on a logarithmic scale. Source: PeopleVsCCIbyAgeHS1_20250115-160715.png

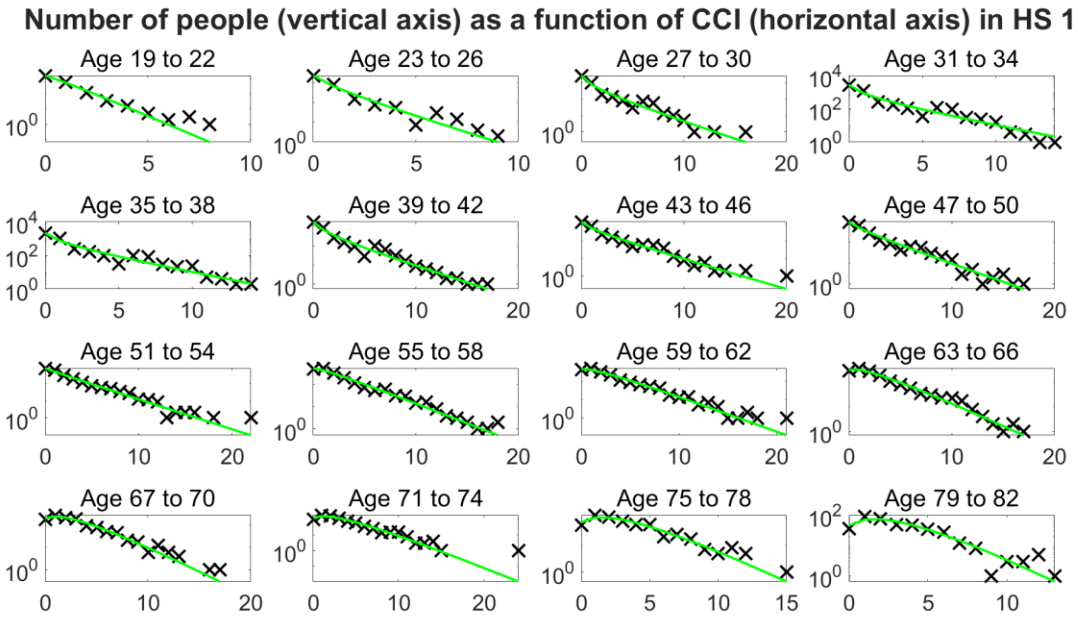
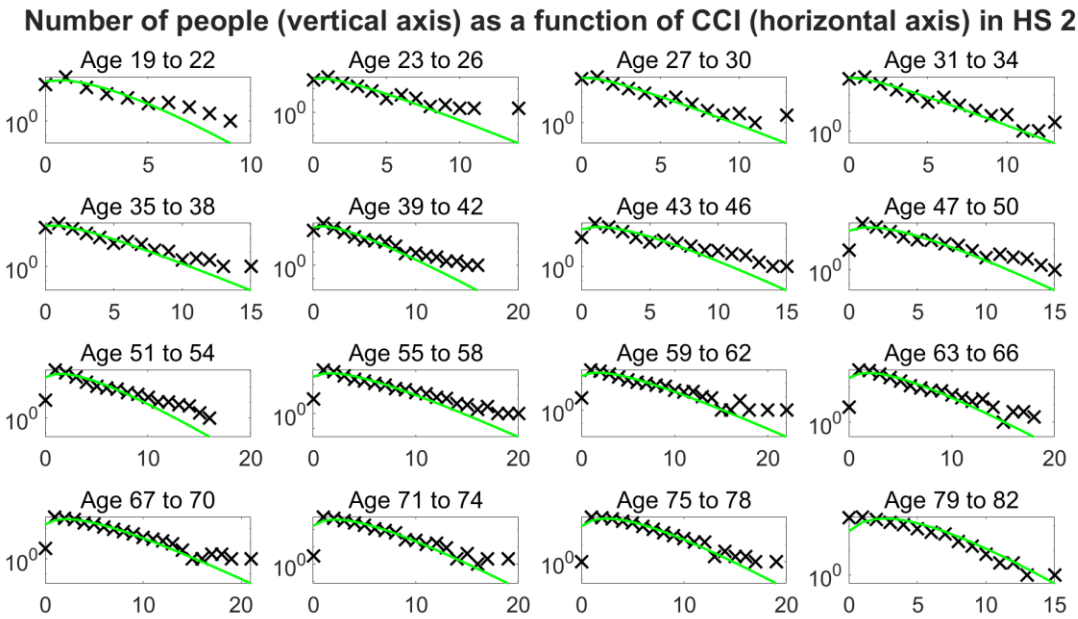


Figure 7 For HS 2, the number of people in each 4-year age group (vertical axis of black ×) who have each value of eCCI (horizontal axis of black ×), and the fitted negative binomial distribution (green line). The sample variance differs most from the variance predicted by the negative binomial distribution in the age groups 36-54. The other health systems do not show similar deviations in these age groups. Source: PeopleVsCCIbyAgeHS2_20250115-160715.png

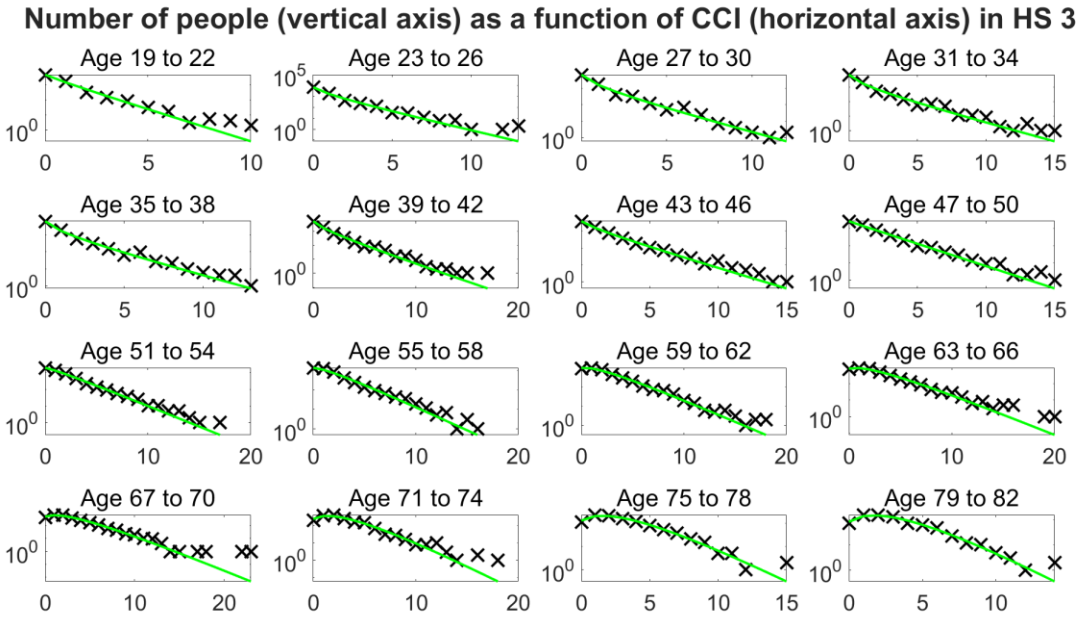


278

279 Figure 8 For HS 3, the number of people in each 4-year age group (vertical axis of black ×) who

280 have each value of eCCI (horizontal axis of black ×), and the fitted negative binomial

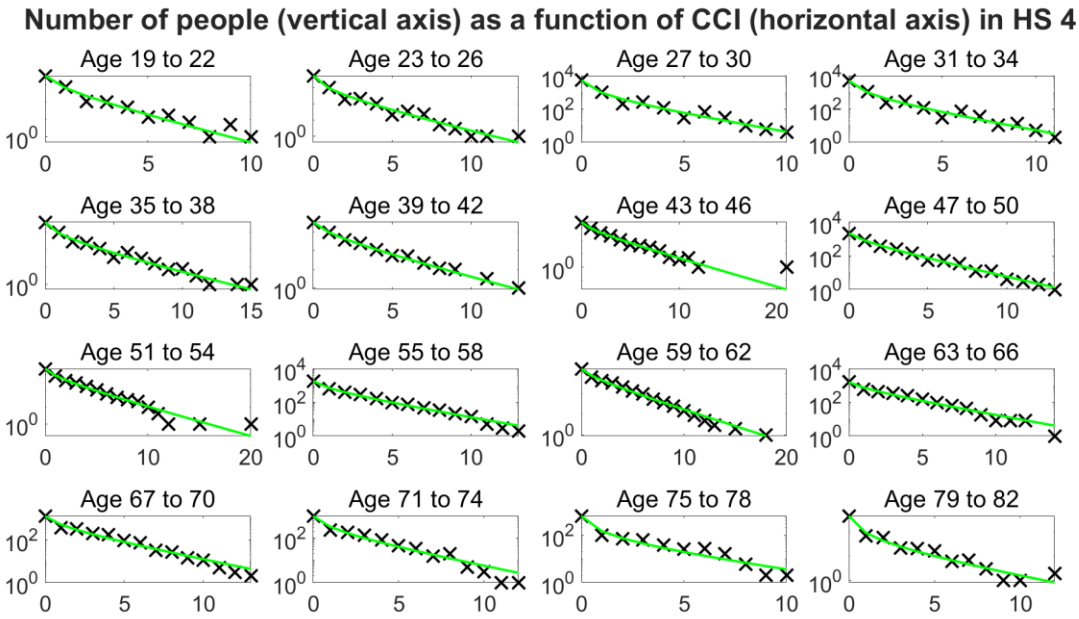
281 distribution (green line). Source: PeopleVsCCIbyAgeHS3_20250115-160715.png



282

283

Figure 9 For HS 4, the number of people in each 4-year age group (vertical axis of black ×) who have each value of eCCI (horizontal axis of black ×), and the fitted negative binomial distribution (green line). Source: PeopleVsCCIbyAgeHS4_20250115-160715.png



Negative binomial distribution

The negative binomial distribution can be parameterized in various ways. The probability density function of the negative binomial distribution in Matlab (https://www.mathworks.com/help/stats/nbinpdf.html), which we used for these computations, is: for $0 < p < 1,\ r > 0$ (r may or may not be an integer),

$$\Pr(eCCI = x) = \left(\frac{\Gamma(r+x)}{\Gamma(r)\Gamma(1+x)}\right) p^r (1-p)^x, \quad x = 0, 1, 2, 3, \ldots \tag{7}$$

If eCCI is negative binomially distributed as in (7), the expectation or population mean of eCCI is

(8) $E(eCCI) = \frac{r(1-p)}{p}$

and the population variance of eCCI is

(9) $Var(eCCI) = \frac{r(1-p)}{p^2}$

(https://www.mathworks.com/help/stats/nbinstat.html). Two possible variance functions for the negative binomial distribution are

(10) $Var(eCCI) = \left(\frac{1}{p}\right) E(eCCI)$ if p is constant and only r varies;

and

(11) $Var(eCCI) = E(eCCI) + \left(\frac{1}{r}\right) [E(eCCI)]^2$ if r is constant and only p varies.

Both (10) and (11) imply $Var(eCCI) > E(eCCI)$.

If (10) holds for every age group a with the same value of p for all age groups and only r varies, then (10) is identical to TL (1) with $c = 1/p$ and $b = 1$. In the limit as p approaches 1, the variance function (10) approaches $Var(eCCI) = E(eCCI)$, which is the variance function of the Poisson distribution, where variance equals mean. If p varies notably with age a, then (1) and (10) are not consistent.

If only p varies and r is constant, then the variance is a quadratic function (11) of the mean. In the limit as r approaches infinity, the variance function (11) again approaches $Var(eCCI) = E(eCCI)$, the variance function of the Poisson distribution. In the limit as r approaches 0, the second term in (11) dominates the first and (11) is asymptotic to TL (1) with $c = 1/r$ and $b = 2$. For fixed finite r, when only p varies, $Var(eCCI)$ is a convex increasing function (11) of $E(eCCI)$, not consistent with TL (1).

Figure 10 plots the estimated negative binomial parameters r, p as a function of age group a for each HS. In all four HSs, with increasing age, r changes more than p. The major trend is that r decreases from age group 20 to 28 and increases from age group 28 onward (except in HS 4 above age 64). In HS 1 and HS 2 (New York), the estimated r (blue line, blue x) notably exceeds the estimated p (orange line, orange +). As age increases, in HSs 1, 2, and 3, p decreases slightly at the earlier ages and then smoothly levels off, while in HS 4, p falls over the entire range of ages (apart from a slight rise between ages 40 and 48).

The trends over age in the estimated parameters of the negative binomial distribution of CCI account qualitatively for the occurrence of Taylor's law with exponent near or less than $b = 1$ in these data. Using the estimated parameter values r, p, Figure 11 plots the population variance $Var(CCI) = r(1-p)/p^2$ as a function of the predicted mean $E(CCI) = r(1-p)/p$ for each age group in each HS (blue x marks) and a (black) curve, a power function corresponding to Taylor's law, fitted by least squares to the 16 values of $(r(1-p)/p, r(1-p)/p^2)$ for the 16 age groups in each HS. Only for HS 2 is $b > 1$ significantly. In HS 1 and 3, $b < 1$ significantly, and b is not significantly different from one in HS 4.

For HS 1 (New York) and HS 3 (Chicago) in Figure 11, $Var(CCI)$ increases as a concave function of the mean $E(CCI)$ compared to the fitted power-law curve. This observation is consistent with the curved path of the circles in Figure 3's top left panel for HS 1 (New York). Likewise, New York's HS 2 gives concordant M-shaped circles for $Var(CCI)$ in Figure 3 and blue x markers in Figure 10, top right panels, compared to the fitted power-law curves. Similarly, the panels for Chicago in the bottom row of Figure 10, like the panels for Chicago in the bottom row of Figure 3, roughly support Taylor's power law variance function.

Figure 10 Estimates of the negative binomial parameters r, p as a function of age group a for each health system. Source: NegativeBinomialParameters_20250115-185504.png

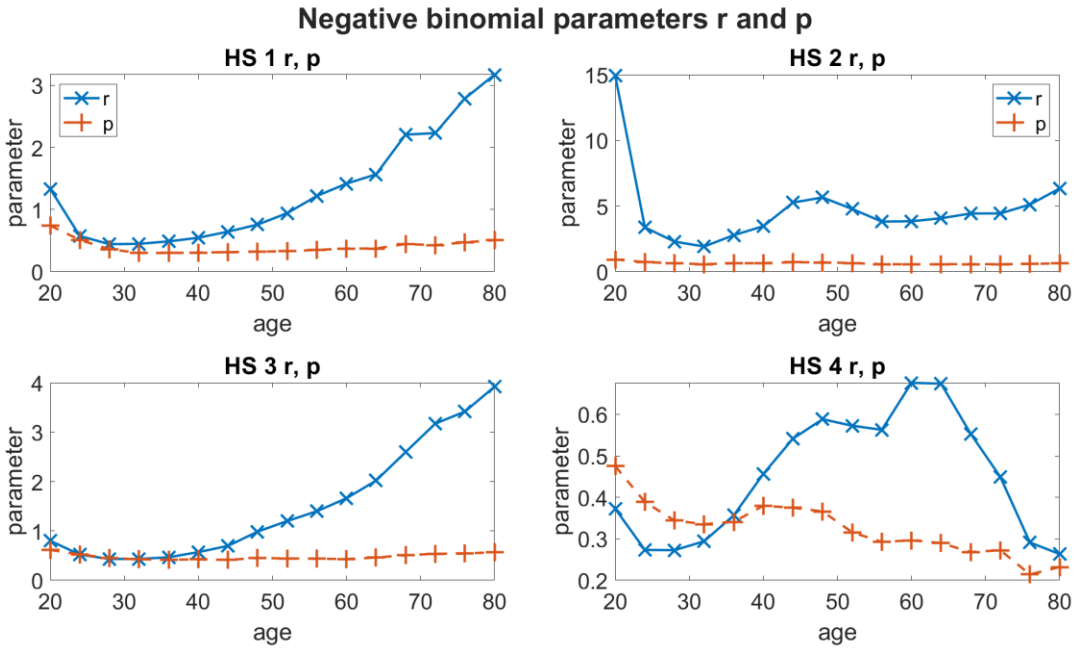
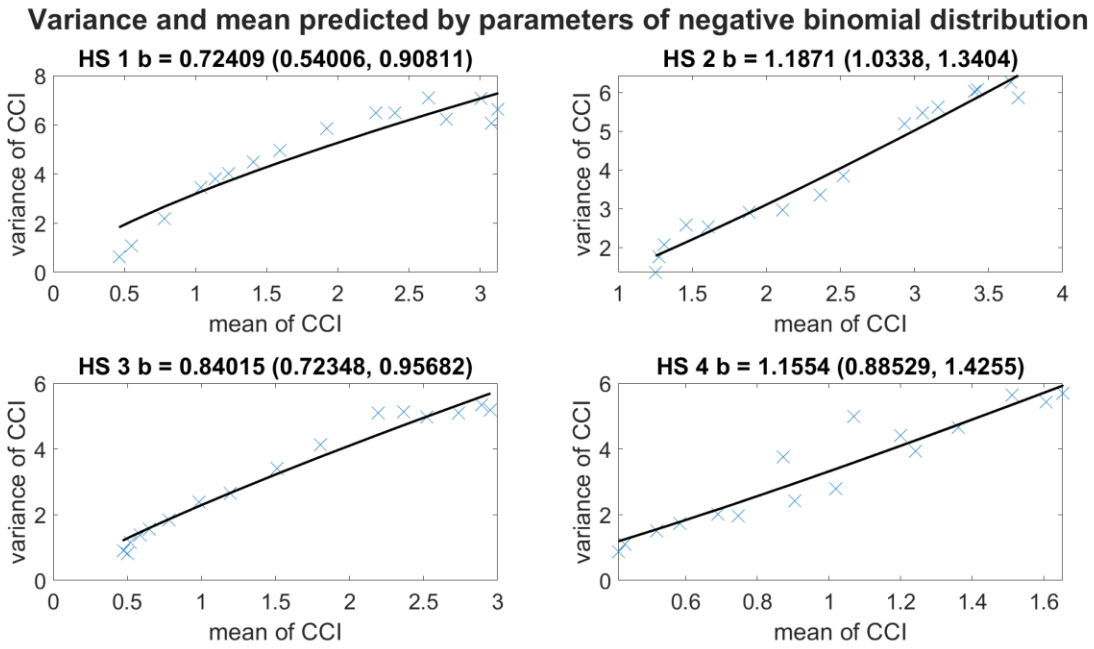


Figure 11 The variance $Var(eCCI) = r(1-p)/p^2$ of a negative binomial distribution of individual eCCI as a function of the mean $E(CCI) = r(1-p)/p$ for each age group in each health system (blue x marks) and a (black) curve, corresponding to a fitted Taylor's power law (1). Source: NegativeBinomialPredictedTLbyHS_20250117-115016.png



DISCUSSION

Summary and interpretation of main results

We show that, in four health systems (HSs), two in New York and two in Chicago, the variation in the enhanced Charlson Comorbidity Index (eCCI) (defined by Charlson et al., 2025) among people of a given age increases with age, apart from random variation in small samples. The increasing variation is systematically related to an increase with age in the average eCCI, in qualitative agreement with Taylor's power law of fluctuation scaling or the quadratic Taylor's

law. The quintiles of eCCI diverge increasingly with age, again revealing increasing scatter in eCCI of individuals of each age. The frequency histogram of eCCI within age groups is generally well described by a negative binomial distribution. The parameters of the negative binomial distribution of different age groups account approximately for the appearance of Taylor's law or the quadratic Taylor's law in these four HSs. The underlying mechanisms of the negative binomial distribution and of Taylor's law or its quadratic generalization remain to be determined.

The differences among the four HSs could be due to:

- differences in the patient population,
- differences in the environment of the different HSs,
- differences in the coding practices of different HSs in different cities and states, including differences among clinicians or health insurance companies in their propensity to add comorbid conditions to records of ageing patients who already have multiple comorbid conditions,
- differences in the survival of older adults with larger eCCI if the survivors to older ages experienced high values of eCCI less often than those who died,
- or to interactions of these and other unnamed factors.

In every HS reported here, in every 4-year age group, the variance of eCCI substantially exceeds the mean eCCI. This observation rejects a Poisson model of variation in eCCI, in which the variance of eCCI equals the mean of eCCI. Poisson variation would arise if every cause of an increase by one unit in eCCI were independent of every other cause and had an equal probability of causing an increase in eCCI. The medical conditions that are counted in eCCI are known *not* to be independent and *not* to be of identical probability. The sample has more variation in eCCI

than the Poisson model can describe, suggesting that the factors that contribute to eCCI are positively associated, i.e., they tend to vary together. Moreover, these associations may change with age.

The clear finding that the variance and the mean of eCCI rise with age means that aging brings increasing comorbidity (known already) and increasing diversity of comorbidity (described here). The upper quantiles of the distribution of eCCI at each age rise more rapidly than the lower quantiles and the mean of eCCI. Increasing age brings more comorbidities than would be predicted from the increase in the mean of eCCI alone. Clinical practice should take account of the increasing diversity of comorbidity.

Interpreting the negative binomial distribution

The negative binomial distribution describes very well the frequency distribution (histogram) of eCCI within nearly every 4-year age group in all 4 health systems. The negative binomial distribution has at least three well known interpretations or models (Bailey 1964; Grimmett and Stirzaker 2001): heterogeneity, contagion, and sampling.

Heterogeneity: the negative binomial is a mixture of Poisson distributions with gamma-distributed means. In greater detail, if an age group contains subgroups, and each subgroup has a Poisson distribution of eCCI, and the mean eCCIs of the different subgroups vary according to a gamma distribution, then the whole age group has a negative binomial distribution of eCCI.

The contagion model's key idea is that the more chronic conditions a person has, at any age, the higher the risk that the person will acquire an additional chronic condition. This hypothetical "contagion" (in a statistical sense, not related specifically to infectious diseases) among conditions could be tested empirically with adequate longitudinal data.

Sampling: This interpretation is possible when the parameter *r* is a positive whole number. Suppose that, over time, each person experiences a succession of independent random trials. Suppose a person's trial "succeeds" with probability p and "fails" with probability $1 - p$. To achieve *r* successes, a person has to experience a variable number *x* of failures since there is a fixed positive probability $1 - p$ of failure at every trial. The number of extra trials *x* a person experiences to reach *r* successes is a random quantity that obeys the negative binomial distribution (7). Here the number of failures *x* corresponds to eCCI, the number of comorbid conditions a person acquires until achieving *r* "successes."

The available data are insufficient to select among these possible mechanisms that lead to the negative binomial distribution. The multiple possibilities suggest questions for future research.

Future research

For clinical applications, to identify risk groups and opportunities for patient education and intervention, it would be highly desirable to repeat this analysis after stratifying each age group by sex, marital status, education, behavior (smoking, drug use, alcohol use, exercise, sleep, diet, relationships), socio-economic status, environmental hazards, and perhaps other characteristics. The auxiliary variables associated with individuals of a given age with high and low eCCI may offer clues to possible preventive interventions or behavioral changes.

The 39 conditions included in the enhanced eCCI generate $2^{39} \approx 5.5 \times 10^{11}$ (almost 550 billion) possible combinations of conditions. Identifying each patient's combination of conditions would make it possible to examine the marginal frequency of each condition and the associations, positive and negative, among conditions.

The present analysis is cross-sectional, comparing individuals who are observed once only at different ages. Longitudinal records of medical history for the oldest age groups would make it possible to calculate the eCCI retrospectively for those individuals, say, at each age 10, 20, 30, … years earlier, and to compare the trajectories (frequency histogram, mean, variance, and relation of variance to mean) over increasing age of individuals at an advanced age who have low eCCI with the trajectories of individuals who have high eCCI. Longitudinal records of medical history would also make it possible to evaluate the influence of selective or differential survival on the distribution of eCCI among the living, and easier to discriminate among alternative mechanisms that lead to the observed negative binomial distribution of eCCI within age groups.

ACKNOWLEDGMENTS

This collaboration among The Rockefeller University Center for Clinical and Translational Science, Clinical Directors Network (CDN), and Weill Cornell Medicine, is part of a cluster randomized clinical trial funded by the Patient Centered Outcomes Research Institute (PCORI).

The trial was funded by Patient-Centered Outcomes Research Institute (PCORI) Contract Number IHS-2017C3-8923 to Clinical Directors Network (CDN) with additional infrastructure support provided by AHRQ (N^2-PBRN Grant #1 P30-HS-021667 to Clinical Directors Network); *Grant* # UL1 TR001866 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program to The Rockefeller University; INSIGHT (PCORI Award #R-1306-03961 to Weill Cornell Medicine); and CAPriCORN (PCORI Award #CDRN-1306-04737 to Northwestern Medicine). PCORI scientific staff played no role in study design and conduct. The

statements presented in this publication are solely the responsibility of the author(s) and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute® (PCORI®), its Board of Governors or Methodology Committee.

J.E.C. thanks Roseanne Benjamin for help during this work and Dr. Ruth Kevess-Cohen M.D. for comments based on her geriatric clinical practice, as well as Mary E. Charlson MD, Martin T. Wells, PhD, James P. Hollenberg, MD, Rosio Ramos, BA, Robin Andrews, DSW, Shelly S. Sital, MPH, and Ilana Mittleman, MPH.

REFERENCES

Bailey, N.T.J. 1964 The elements of stochastic processes with applications to the natural sciences. Wiley, New York.

John R Beard, Alana Officer, Islene Araujo de Carvalho, Ritu Sadana, Anne Margriet Pot, Jean-Pierre Michel, Peter Lloyd-Sherlock, JoAnne E Epping-Jordan, G M E E (Geeske) Peeters, Wahyu Retno Mahanani, Jotheeswaran Amuthavalli Thiyagarajan, Somnath Chatterji 2016 The World report on ageing and health: a policy framework for healthy ageing. Lancet 387:2145-54.

Bensken WP, Navale SM, McGrath BM, et al. Variation in multimorbidity by sociodemographics and social drivers of health among patients seen at community-based health centers. Journal of Multimorbidity and Comorbidity. 2024;14. doi:10.1177/26335565241236410

Charlson, Mary, Szatrowski, Ted P., Peterson, Janey, and Gold, Jeffrey 1994 Validation of a combined comorbidity index. Journal of Clinical Epidemiology 47(11):1245-1251 (April).

Charlson, Mary E., Carrozzino, Danilo, Guidi, Jenny, and Patierno, Chiara 2022 Charlson Comorbidity Index: A critical review of clinimetric properties. Psychotherapy and Psychosomatics 91:8-35. DOI: 10.1159/000521288

Charlson Mary E. and Martin T. Wells 2022 Comorbidity: from a confounder in longitudinal clinical research to the main issue in population management. Psychotherapy and Psychosomatics 91:145-151 DOI: 10.1159/000521952

Charlson ME, Mittleman I, Ramos R, Cassells A, Lin TJ, Eggleston A, Wells MT, Hollenberg J, Pirraglia P, Winston G, Tobin JN. Preventing "tipping points" in high comorbidity patients: A lifeline from health coaches - rationale, design and methods. Contemp Clin Trials. 2025 Feb 28:107865. doi: 10.1016/j.cct.2025.107865. Epub ahead of print. PMID: 40024364.

Chen Y, Shi L, Zheng X, Yang J, Xue Y, Xiao S, Xue B, Zhang J, Li X, Lin H, et al. Patterns and determinants of multimorbidity in older adults: study in health-ecological perspective. International Journal of Environmental Research and Public Health. 2022; 19(24):16756. https://doi.org/10.3390/ijerph192416756

Chen, HL., Yu, XH., Yin, YH. et al. Multimorbidity patterns and the association with health status of the oldest-old in long-term care facilities in China: a two-step analysis. BMC Geriatr 23, 851 (2023). https://doi.org/10.1186/s12877-023-04507-8

Saifur Rahman Chowdhury, Dipak Chandra Das, Tachlima Chowdhury Sunna, Joseph Beyene, Ahmed Hossain, Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. eClinicalMedicine Open Access Published:February 15, 2023 DOI:https://doi.org/10.1016/j.eclinm.2023.101860

Miguel J. Divo, Carlos H. Martinez, David M. Mannino, Ageing and the epidemiology of multimorbidity. European Respiratory Journal Oct 2014, 44 (4) 1055-1068; DOI: 10.1183/09031936.00059814

Grimmett, Geoffrey R. and Stirzaker, David R. 2001 Probability and random processes, 3rd Edition. Oxford University Press, New York.

Hultsch DF, MacDonald SW, Dixon RA. Variability in reaction time performance of younger and older adults. J Gerontol B Psychol Sci Soc Sci. 2002 Mar;57(2):P101-15. doi: 10.1093/geronb/57.2.p101. PMID: 11867658.

Ly, Kevin, Wakefield, Dorothy, and ZuWallack, Richard 2025 The usefulness of Charlson Comorbidity Index (CCI) scoring in predicting all-cause mortality in outpatients with clinical diagnoses of COPD. Journal of Multimorbidity and Comorbidity 15:1-7. doi:10.1177/26335565251315876

Santoni G, Angleman S, Welmer A-K, Mangialasche F, Marengoni A, Fratiglioni L 2015 Age-related variation in health status after age 60. PLoS ONE 10(3):e0120077. https://doi.org/10.1371/journal.pone.0120077

Taylor, L. R. 1961 Aggregation, variance and the mean. Nature 189(4766):732 735, March 4. doi:10.1038/189732a0

Taylor R. A. J. (2019) *Taylor's power law: order and pattern in nature.* Academic Press: London, UK.

Taylor, L. R., Woiwod, I. P. and Perry, J. N. 1978 The density-dependence of spatial behaviour and the rarity of randomness. Journal of Animal Ecology 47(2):383 406, June. DOI 10.2307/3790. Stable URL: http://www.jstor.org/stable/3790

506 Taylor, L. R., Perry, J. N., Woiwod, I. P. and Taylor, R. A. J. 1988 Specificity of the spatial
507 power-law exponent in ecology and agriculture. Nature 332(21):721-722. DOI
508 10.1038/332721a0