

Measuring the quality of mortality data in high-income settings

Amy Mann, Mathew Kiang, Monica Alexander

Toronto Data Science Workshop, January 14, 2026



Motivation

Motivation

Motivation



[NASEM; 2021]

Motivation



[NASEM; 2021]

- More than 5,000 papers have cited US CDC's vital statistics system in the last ten years

Motivation



[NASEM; 2021]

- More than 5,000 papers have cited US CDC's vital statistics system in the last ten years
- Often taken for granted in high-income countries that mortality data is reliable

Motivation



[NASEM; 2021]

- More than 5,000 papers have cited US CDC's vital statistics system in the last ten years
- Often taken for granted in high-income countries that mortality data is reliable
- Most data quality papers focus on low-income settings and national registries. We are interested in sub-national differences in data quality

Death certificate coding

Death certificate coding

U.S. STANDARD CERTIFICATE OF DEATH									
DEATH CERTIFICATE REGISTRATION AND REPORT FORM FOR MATERIALS NOT OTHERWISE CODED BY STATE OR LOCAL HEALTH DEPARTMENT DECEASED		DECEASED NAME (First Middle Last)		DATE OF DEATH (Month/Day/Year)					
4. SOCIAL SECURITY NUMBER Mr./Ms./Other Social Security Number Date of Birth		5. UNITS 1 YEAR Mo. Weeks Days Hours Minutes		6. DATE OF BIRTH (Month/Day/Year)		7. BIRTHPLACE (City and State or Foreign Country)			
8. WAS DECEASED EVER IN U.S. ARMED FORCES? <input type="checkbox"/> Yes <input type="checkbox"/> No		9. PLACE OF DEATH (Check one) Home Hospital Nursing home Residence Other (Specify)		10. DEATH CERTIFICATION (Check one or more) Physical Mental Death from natural causes Disease Injury Homicide Other (Specify)					
11. FACILITY NAME (If not indicated, give street and number)		12. DEATH TOWN OR LOCATION OF DEATH		13. COUNTY OF DEATH					
14. MARITAL STATUS: Married Divorced Widowed Never married Unknown (Specify)		15. DECEASED'S USUAL OCCUPATION Work, profession, trade, or business (Do not use names)		16. KIND OF BUSINESS/INDUSTRY					
17. RESIDENCE STATE 18. COUNTY 19. CITY, TOWNS, OR LOCALITY 20. STREET AND NUMBER		21. ADDRESS OF DECEASED'S PREVIOUS RESIDENCE State City Town Street Phone Name Signature		22. RACE—American Indian Asian and Pacific Islander Black—African American Hispanic—Latino White—Caucasian Unknown (Specify) (If Hispanic, check 21-A or 21-C)		23. EDUCATION LEVEL Elementary school High school College or above Unknown			
24. FATHER'S NAME (First Middle Last)		25. MOTHER'S NAME (First Middle Maiden Surname)		26. Mailing Address (Street and Number or Rural Route Number, City or Town, State, Zip Code)					
27. INFORMANT'S NAME (First Middle Last)		28. METHOD OF DISPOSITION <input type="checkbox"/> Burial <input type="checkbox"/> Cremation <input type="checkbox"/> Removal from State <input type="checkbox"/> Removal from Country <input type="checkbox"/> Burial at Sea		29. PLACE OF DISPOSITION (Name of cemetery, cemetery, or other place)		30. LOCATION—City or Town, State			
31. DEPOSITION BY AUTHORITY ON OTHER SIDE		32. SIGNATURE OF PERSON ACTING AS CORROBORATE FOR DEATH OR PERSON ACTING AS SUCH		33. LICENSE NUMBER of Licensed		34. NAME AND ADDRESS OF FACILITY Person Acting As Such			
35. PRESENTING PHYSICIAN (Name) 36. AMOUNT AND TYPE OF MEDICAL EXAMINATION PERFORMED 37. PRINCIPAL EXAMINER		38. DATE OF DEATH (Month/Day/Year)		39. DATE OF AUTOPSY (Month/Day/Year)		40. DATE BORN (Month/Day/Year)			
35. Complete with 36-a through 36-f, if applicable. If not applicable, check here _____ not applicable at time of death Signature and		38. To the best of my knowledge, death occurred at the time, date, and place stated.		39. DEATH NUMBER (Month/Day/Year)		40. DATE BORN (Month/Day/Year)			
36. TIME OF DEATH M: _____ S: _____ H: _____ A.M. _____ P.M. _____		38. DATE PROBABLE (Probable Month/Year)		39. WAS CASE REFERRED TO MEDICAL EXAMINER/Coroner (Yes or no)		40. DATE BORN (Month/Day/Year)			
37. PART I. Give the medical history, or circumstances, that caused the death. Do not enter the mode of dying, such as cardiac or respiratory arrest, sudden death, or death due to external causes. If more than one cause of death, list them in Part II. IMMEDIATE CAUSE OF DEATH List the immediate cause(s) of death. If more than one cause, list them in Part II. 1. Due to _____ 2. Due to _____ 3. Due to _____ 4. Due to _____ 5. Due to _____ 6. Due to _____ 7. Due to _____ 8. Due to _____ 9. Due to _____ 10. Due to _____ 11. Due to _____ 12. Due to _____ 13. Due to _____ 14. Due to _____ 15. Due to _____ 16. Due to _____ 17. Due to _____ 18. Due to _____ 19. Due to _____ 20. Due to _____ 21. Due to _____ 22. Due to _____ 23. Due to _____ 24. Due to _____ 25. Due to _____ 26. Due to _____ 27. Due to _____ 28. Due to _____ 29. Due to _____ 30. Due to _____ 31. Due to _____ 32. Due to _____ 33. Due to _____ 34. Due to _____ 35. Due to _____ 36. Due to _____ 37. Due to _____ 38. Due to _____ 39. Due to _____ 40. Due to _____ 41. Due to _____ 42. Due to _____ 43. Due to _____ 44. Due to _____ 45. Due to _____ 46. Due to _____ 47. Due to _____ 48. Due to _____ 49. Due to _____ 50. Due to _____ 51. Due to _____ 52. Due to _____ 53. Due to _____ 54. Due to _____ 55. Due to _____ 56. Due to _____ 57. Due to _____ 58. Due to _____ 59. Due to _____ 60. Due to _____ 61. Due to _____ 62. Due to _____ 63. Due to _____ 64. Due to _____ 65. Due to _____ 66. Due to _____ 67. Due to _____ 68. Due to _____ 69. Due to _____ 70. Due to _____ 71. Due to _____ 72. Due to _____ 73. Due to _____ 74. Due to _____ 75. Due to _____ 76. Due to _____ 77. Due to _____ 78. Due to _____ 79. Due to _____ 80. Due to _____ 81. Due to _____ 82. Due to _____ 83. Due to _____ 84. Due to _____ 85. Due to _____ 86. Due to _____ 87. Due to _____ 88. Due to _____ 89. Due to _____ 90. Due to _____ 91. Due to _____ 92. Due to _____ 93. Due to _____ 94. Due to _____ 95. Due to _____ 96. Due to _____ 97. Due to _____ 98. Due to _____ 99. Due to _____ 100. Due to _____ 101. Due to _____ 102. Due to _____ 103. Due to _____ 104. Due to _____ 105. Due to _____ 106. Due to _____ 107. Due to _____ 108. Due to _____ 109. Due to _____ 110. Due to _____ 111. Due to _____ 112. Due to _____ 113. Due to _____ 114. Due to _____ 115. Due to _____ 116. Due to _____ 117. Due to _____ 118. Due to _____ 119. Due to _____ 120. Due to _____ 121. Due to _____ 122. Due to _____ 123. Due to _____ 124. Due to _____ 125. Due to _____ 126. Due to _____ 127. Due to _____ 128. Due to _____ 129. Due to _____ 130. Due to _____ 131. Due to _____ 132. Due to _____ 133. Due to _____ 134. Due to _____ 135. Due to _____ 136. Due to _____ 137. Due to _____ 138. Due to _____ 139. Due to _____ 140. Due to _____ 141. Due to _____ 142. Due to _____ 143. Due to _____ 144. Due to _____ 145. Due to _____ 146. Due to _____ 147. Due to _____ 148. Due to _____ 149. Due to _____ 150. Due to _____ 151. Due to _____ 152. Due to _____ 153. Due to _____ 154. Due to _____ 155. Due to _____ 156. Due to _____ 157. Due to _____ 158. Due to _____ 159. Due to _____ 160. Due to _____ 161. Due to _____ 162. Due to _____ 163. Due to _____ 164. Due to _____ 165. Due to _____ 166. Due to _____ 167. Due to _____ 168. Due to _____ 169. Due to _____ 170. Due to _____ 171. Due to _____ 172. Due to _____ 173. Due to _____ 174. Due to _____ 175. Due to _____ 176. Due to _____ 177. Due to _____ 178. Due to _____ 179. Due to _____ 180. Due to _____ 181. Due to _____ 182. Due to _____ 183. Due to _____ 184. Due to _____ 185. Due to _____ 186. Due to _____ 187. Due to _____ 188. Due to _____ 189. Due to _____ 190. Due to _____ 191. Due to _____ 192. Due to _____ 193. Due to _____ 194. Due to _____ 195. Due to _____ 196. Due to _____ 197. Due to _____ 198. Due to _____ 199. Due to _____ 200. Due to _____ 201. Due to _____ 202. Due to _____ 203. Due to _____ 204. Due to _____ 205. Due to _____ 206. Due to _____ 207. Due to _____ 208. Due to _____ 209. Due to _____ 210. Due to _____ 211. Due to _____ 212. Due to _____ 213. Due to _____ 214. Due to _____ 215. Due to _____ 216. Due to _____ 217. Due to _____ 218. Due to _____ 219. Due to _____ 220. Due to _____ 221. Due to _____ 222. Due to _____ 223. Due to _____ 224. Due to _____ 225. Due to _____ 226. Due to _____ 227. Due to _____ 228. Due to _____ 229. Due to _____ 230. Due to _____ 231. Due to _____ 232. Due to _____ 233. Due to _____ 234. Due to _____ 235. Due to _____ 236. Due to _____ 237. Due to _____ 238. Due to _____ 239. Due to _____ 240. Due to _____ 241. Due to _____ 242. Due to _____ 243. Due to _____ 244. Due to _____ 245. Due to _____ 246. Due to _____ 247. Due to _____ 248. Due to _____ 249. Due to _____ 250. Due to _____ 251. Due to _____ 252. Due to _____ 253. Due to _____ 254. Due to _____ 255. Due to _____ 256. Due to _____ 257. Due to _____ 258. Due to _____ 259. Due to _____ 260. Due to _____ 261. Due to _____ 262. Due to _____ 263. Due to _____ 264. Due to _____ 265. Due to _____ 266. Due to _____ 267. Due to _____ 268. Due to _____ 269. Due to _____ 270. Due to _____ 271. Due to _____ 272. Due to _____ 273. Due to _____ 274. Due to _____ 275. Due to _____ 276. Due to _____ 277. Due to _____ 278. Due to _____ 279. Due to _____ 280. Due to _____ 281. Due to _____ 282. Due to _____ 283. Due to _____ 284. Due to _____ 285. Due to _____ 286. Due to _____ 287. Due to _____ 288. Due to _____ 289. Due to _____ 290. Due to _____ 291. Due to _____ 292. Due to _____ 293. Due to _____ 294. Due to _____ 295. Due to _____ 296. Due to _____ 297. Due to _____ 298. Due to _____ 299. Due to _____ 300. Due to _____ 301. Due to _____ 302. Due to _____ 303. Due to _____ 304. Due to _____ 305. Due to _____ 306. Due to _____ 307. Due to _____ 308. Due to _____ 309. Due to _____ 310. Due to _____ 311. Due to _____ 312. Due to _____ 313. Due to _____ 314. Due to _____ 315. Due to _____ 316. Due to _____ 317. Due to _____ 318. Due to _____ 319. Due to _____ 320. Due to _____ 321. Due to _____ 322. Due to _____ 323. Due to _____ 324. Due to _____ 325. Due to _____ 326. Due to _____ 327. Due to _____ 328. Due to _____ 329. Due to _____ 330. Due to _____ 331. Due to _____ 332. Due to _____ 333. Due to _____ 334. Due to _____ 335. Due to _____ 336. Due to _____ 337. Due to _____ 338. Due to _____ 339. Due to _____ 340. Due to _____ 341. Due to _____ 342. Due to _____ 343. Due to _____ 344. Due to _____ 345. Due to _____ 346. Due to _____ 347. Due to _____ 348. Due to _____ 349. Due to _____ 350. Due to _____ 351. Due to _____ 352. Due to _____ 353. Due to _____ 354. Due to _____ 355. Due to _____ 356. Due to _____ 357. Due to _____ 358. Due to _____ 359. Due to _____ 360. Due to _____ 361. Due to _____ 362. Due to _____ 363. Due to _____ 364. Due to _____ 365. Due to _____ 366. Due to _____ 367. Due to _____ 368. Due to _____ 369. Due to _____ 370. Due to _____ 371. Due to _____ 372. Due to _____ 373. Due to _____ 374. Due to _____ 375. Due to _____ 376. Due to _____ 377. Due to _____ 378. Due to _____ 379. Due to _____ 380. Due to _____ 381. Due to _____ 382. Due to _____ 383. Due to _____ 384. Due to _____ 385. Due to _____ 386. Due to _____ 387. Due to _____ 388. Due to _____ 389. Due to _____ 390. Due to _____ 391. Due to _____ 392. Due to _____ 393. Due to _____ 394. Due to _____ 395. Due to _____ 396. Due to _____ 397. Due to _____ 398. Due to _____ 399. Due to _____ 400. Due to _____ 401. Due to _____ 402. Due to _____ 403. Due to _____ 404. Due to _____ 405. Due to _____ 406. Due to _____ 407. Due to _____ 408. Due to _____ 409. Due to _____ 410. Due to _____ 411. Due to _____ 412. Due to _____ 413. Due to _____ 414. Due to _____ 415. Due to _____ 416. Due to _____ 417. Due to _____ 418. Due to _____ 419. Due to _____ 420. Due to _____ 421. Due to _____ 422. Due to _____ 423. Due to _____ 424. Due to _____ 425. Due to _____ 426. Due to _____ 427. Due to _____ 428. Due to _____ 429. Due to _____ 430. Due to _____ 431. Due to _____ 432. Due to _____ 433. Due to _____ 434. Due to _____ 435. Due to _____ 436. Due to _____ 437. Due to _____ 438. Due to _____ 439. Due to _____ 440. Due to _____ 441. Due to _____ 442. Due to _____ 443. Due to _____ 444. Due to _____ 445. Due to _____ 446. Due to _____ 447. Due to _____ 448. Due to _____ 449. Due to _____ 450. Due to _____ 451. Due to _____ 452. Due to _____ 453. Due to _____ 454. Due to _____ 455. Due to _____ 456. Due to _____ 457. Due to _____ 458. Due to _____ 459. Due to _____ 460. Due to _____ 461. Due to _____ 462. Due to _____ 463. Due to _____ 464. Due to _____ 465. Due to _____ 466. Due to _____ 467. Due to _____ 468. Due to _____ 469. Due to _____ 470. Due to _____ 471. Due to _____ 472. Due to _____ 473. Due to _____ 474. Due to _____ 475. Due to _____ 476. Due to _____ 477. Due to _____ 478. Due to _____ 479. Due to _____ 480. Due to _____ 481. Due to _____ 482. Due to _____ 483. Due to _____ 484. Due to _____ 485. Due to _____ 486. Due to _____ 487. Due to _____ 488. Due to _____ 489. Due to _____ 490. Due to _____ 491. Due to _____ 492. Due to _____ 493. Due to _____ 494. Due to _____ 495. Due to _____ 496. Due to _____ 497. Due to _____ 498. Due to _____ 499. Due to _____ 500. Due to _____ 501. Due to _____ 502. Due to _____ 503. Due to _____ 504. Due to _____ 505. Due to _____ 506. Due to _____ 507. Due to _____ 508. Due to _____ 509. Due to _____ 510. Due to _____ 511. Due to _____ 512. Due to _____ 513. Due to _____ 514. Due to _____ 515. Due to _____ 516. Due to _____ 517. Due to _____ 518. Due to _____ 519. Due to _____ 520. Due to _____ 521. Due to _____ 522. Due to _____ 523. Due to _____ 524. Due to _____ 525. Due to _____ 526. Due to _____ 527. Due to _____ 528. Due to _____ 529. Due to _____ 530. Due to _____ 531. Due to _____ 532. Due to _____ 533. Due to _____ 534. Due to _____ 535. Due to _____ 536. Due to _____ 537. Due to _____ 538. Due to _____ 539. Due to _____ 540. Due to _____ 541. Due to _____ 542. Due to _____ 543. Due to _____ 544. Due to _____ 545. Due to _____ 546. Due to _____ 547. Due to _____ 548. Due to _____ 549. Due to _____ 550. Due to _____ 551. Due to _____ 552. Due to _____ 553. Due to _____ 554. Due to _____ 555. Due to _____ 556. Due to _____ 557. Due to _____ 558. Due to _____ 559. Due to _____ 560. Due to _____ 561. Due to _____ 562. Due to _____ 563. Due to _____ 564. Due to _____ 565. Due to _____ 566. Due to _____ 567. Due to _____ 568. Due to _____ 569. Due to _____ 570. Due to _____ 571. Due to _____ 572. Due to _____ 573. Due to _____ 574. Due to _____ 575. Due to _____ 576. Due to _____ 577. Due to _____ 578. Due to _____ 579. Due to _____ 580. Due to _____ 581. Due to _____ 582. Due to _____ 583. Due to _____ 584. Due to _____ 585. Due to _____ 586. Due to _____ 587. Due to _____ 588. Due to _____ 589. Due to _____ 590. Due to _____ 591. Due to _____ 592. Due to _____ 593. Due to _____ 594. Due to _____ 595. Due to _____ 596. Due to _____ 597. Due to _____ 598. Due to _____ 599. Due to _____ 600. Due to _____ 601. Due to _____ 602. Due to _____ 603. Due to _____ 604. Due to _____ 605. Due to _____ 606. Due to _____ 607. Due to _____ 608. Due to _____ 609. Due to _____ 610. Due to _____ 611. Due to _____ 612. Due to _____ 613. Due to _____ 614. Due to _____ 615. Due to _____ 616. Due to _____ 617. Due to _____ 618. Due to _____ 619. Due to _____ 620. Due to _____ 621. Due to _____ 622. Due to _____ 623. Due to _____ 624. Due to _____ 625. Due to _____ 626. Due to _____ 627. Due to _____ 628. Due to _____ 629. Due to _____ 630. Due to _____ 631. Due to _____ 632. Due to _____ 633. Due to _____ 634. Due to _____ 635. Due to _____ 636. Due to _____ 637. Due to _____ 638. Due to _____ 639. Due to _____ 640. Due to _____ 641. Due to _____ 642. Due to _____ 643. Due to _____ 644. Due to _____ 645. Due to _____ 646. Due to _____ 647. Due to _____ 648. Due to _____ 649. Due to _____ 650. Due to _____ 651. Due to _____ 652. Due to _____ 653. Due to _____ 654. Due to _____ 655. Due to _____ 656. Due to _____ 657. Due to _____ 658. Due to _____ 659. Due to _____ 660. Due to _____ 661. Due to _____ 662. Due to _____ 663. Due to _____ 664. Due to _____ 665. Due to _____ 666. Due to _____ 667. Due to _____ 668. Due to _____ 669. Due to _____ 670. Due to _____ 671. Due to _____ 672. Due to _____ 673. Due to _____ 674. Due to _____ 675. Due to _____ 676. Due to _____ 677. Due to _____ 678. Due to _____ 679. Due to _____ 680. Due to _____ 681. Due to _____ 682. Due to _____ 683. Due to _____ 684. Due to _____ 685. Due to _____ 686. Due to _____ 687. Due to _____ 688. Due to _____ 689. Due to _____ 690. Due to _____ 691. Due to _____ 692. Due to _____ 693. Due to _____ 694. Due to _____ 695. Due to _____ 696. Due to _____ 697. Due to _____ 698. Due to _____ 699. Due to _____ 700. Due to _____ 701. Due to _____ 702. Due to _____ 703. Due to _____ 704. Due to _____ 705. Due to _____ 706. Due to _____ 707. Due to _____ 708. Due to _____ 709. Due to _____ 710. Due to _____ 711. Due to _____ 712. Due to _____ 713. Due to _____ 714. Due to _____ 715. Due to _____ 716. Due to _____ 717. Due to _____ 718. Due to _____ 719. Due to _____ 720. Due to _____ 721. Due to _____ 722. Due to _____ 723. Due to _____ 724. Due to _____ 725. Due to _____ 726. Due to _____ 727. Due to _____ 728. Due to _____ 729. Due to _____ 730. Due to _____ 731. Due to _____ 732. Due to _____ 733. Due to _____ 734. Due to _____ 735. Due to _____ 736. Due to _____ 737. Due to _____ 738. Due to _____ 739. Due to _____ 740. Due to _____ 741. Due to _____ 742. Due to _____ 743. Due to _____ 744. Due to _____ 745. Due to _____ 746. Due to _____ 747. Due to _____ 748. Due to _____ 749. Due to _____ 750. Due to _____ 751. Due to _____ 752. Due to _____ 753. Due to _____ 754. Due to _____ 755. Due to _____ 756. Due to _____ 757. Due to _____ 758. Due to _____ 759. Due to _____ 760. Due to _____ 761. Due to _____ 762. Due to _____ 763. Due to _____ 764. Due to _____ 765. Due to _____ 766. Due to _____ 767. Due to _____ 768. Due to _____ 769. Due to _____ 770. Due to _____ 771. Due to _____ 772. Due to _____ 773. Due to _____ 774. Due to _____ 775. Due to _____ 776. Due to _____ 777. Due to _____ 778. Due to _____ 779. Due to _____ 780. Due to _____ 781. Due to _____ 782. Due to _____ 783. Due to _____ 784. Due to _____ 785. Due to _____ 786. Due to _____ 787. Due to _____ 788. Due to _____ 789. Due to _____ 790. Due to _____ 791. Due to _____ 792. Due to _____ 793. Due to _____ 794. Due to _____ 795. Due to _____ 796. Due to _____ 797. Due to _____ 798. Due to _____ 799. Due to _____ 800. Due to _____ 801. Due to _____ 802. Due to _____ 803. Due to _____ 804. Due to _____ 805. Due to _____ 806. Due to _____ 807. Due to _____ 808. Due to _____ 809. Due to _____ 810. Due to _____ 811. Due to _____ 812. Due to _____ 813. Due to _____ 814. Due to _____ 815. Due to _____ 816. Due to _____ 817. Due to _____ 818. Due to _____ 819. Due to _____ 820. Due to _____ 821. Due to _____ 822. Due to _____ 823. Due to _____ 824. Due to _____ 825. Due to _____ 826. Due to _____ 827. Due to _____ 828. Due to _____ 829. Due to _____ 830. Due to _____ 831. Due to _____ 832. Due to _____ 833. Due to _____ 834. Due to _____ 835. Due to _____ 836. Due to _____ 837. Due to _____ 838. Due to _____ 839. Due to _____ 840. Due to _____ 841. Due to _____ 842. Due to _____ 843. Due to _____ 844. Due to _____ 845. Due to _____ 846. Due to _____ 847. Due to _____ 848. Due to _____ 849. Due to _____ 850. Due to _____ 851. Due to _____ 852. Due to _____ 853. Due to _____ 854. Due to _____ 855. Due to _____ 856. Due to _____ 857. Due to _____ 858. Due to _____ 859. Due to _____ 860. Due to _____ 861. Due to _____ 862. Due to _____ 863. Due to _____ 864. Due to _____ 865. Due to _____ 866. Due to _____ 867. Due to _____ 868. Due to _____ 869. Due to _____ 870. Due to _____ 871. Due to _____ 872. Due to _____ 873. Due to _____ 874. Due to _____ 875. Due to _____ 876. Due to _____ 877. Due to _____ 878. Due to _____ 879. Due to _____ 880. Due to _____ 881. Due to _____ 882. Due to _____ 883. Due to _____ 884. Due to _____ 885. Due to _____ 886. Due to _____ 887. Due to _____ 888. Due to _____ 889. Due to _____ 890. Due to _____ 891. Due to _____ 892. Due to _____ 893. Due to _____ 894. Due to _____ 895. Due to _____ 896. Due to _____ 897. Due to _____ 898. Due to _____ 899. Due to _____ 900. Due to _____ 901. Due to _____ 902. Due to _____ 903. Due to _____ 904. Due to _____ 905. Due to _____ 906. Due to _____ 907. Due to _____ 908. Due to _____ 909. Due to _____ 910. Due to _____ 911. Due to _____ 912. Due to _____ 913. Due to _____ 914. Due to _____ 915. Due to _____ 916. Due to _____ 917. Due to _____ 918. Due to _____ 919. Due to _____ 920. Due to _____ 921. Due to _____ 922. Due to _____ 923. Due to _____ 924. Due to _____ 925. Due to _____ 926. Due to _____ 927. Due to _____ 928. Due to _____ 929. Due to _____ 930. Due to _____ 931. Due to _____ 932. Due to _____ 933. Due to _____ 934. Due to _____ 935. Due to _____ 936. Due to _____ 937. Due to _____ 938. Due to _____ 939. Due to _____ 940. Due to _____ 941. Due to _____ 942. Due to _____ 943. Due to _____ 944. Due to _____ 945. Due to _____ 946. Due to _____ 947. Due to _____ 948. Due to _____ 949. Due to _____ 950. Due to _____ 951. Due to _____ 952. Due to _____ 953. Due to _____ 954. Due to _____ 955. Due to _____ 956. Due to _____ 957. Due to _____ 958. Due to _____ 959. Due to _____ 960. Due to _____ 961. Due to _____ 962. Due to _____ 963. Due to _____ 964. Due to _____ 965. Due to _____ 966. Due to _____ 967. Due to _____ 968. Due to _____ 969. Due to _____ 970. Due to _____ 971. Due to _____ 972. Due to _____ 973. Due to _____ 974. Due to _____ 975. Due to _____ 976. Due to _____ 977. Due to _____ 978. Due to _____ 979. Due to _____ 980. Due to _____ 981. Due to _____ 982. Due to _____ 983. Due to _____ 984. Due to _____ 985. Due to _____ 986. Due to _____ 987. Due to _____ 988. Due to _____ 989. Due to _____ 990. Due to _____ 991. Due to _____ 992. Due to _____ 993. Due to _____ 994. Due to _____ 995. Due to _____ 996. Due to _____ 997. Due to _____ 998. Due to _____ 999. Due to _____ 1000. Due to _____ 1001. Due to _____ 1002. Due to _____ 1003. Due to _____ 1004. Due to _____ 1005. Due to _____ 1006. Due to _____ 1007. Due to _____ 1008. Due to _____ 1009. Due to _____ 1010. Due to _____ 1011. Due to _____ 1012. Due to _____ 1013. Due to _____ 1014. Due to _____ 1015. Due to _____ 1016. Due to _____ 1017. Due to _____ 1018. Due to _____ 1019. Due to _____ 1020. Due to _____ 1021. Due to _____ 1022. Due to _____ 1023. Due to _____ 1024. Due to _____ 1025. Due to _____ 1026. Due to _____ 1027. Due to _____ 1028. Due to _____ 1029. Due to _____ 1030. Due to _____ 1031. Due to _____ 1032. Due to _____ 1033. Due to _____ 1034. Due to _____ 1035. Due to _____ 1036. Due to _____ 1037. Due to _____ 1038. Due to _____ 1039. Due to _____ 1040. Due to _____ 1041. Due to _____ 1042. Due to _____ 1043. Due to _____ 1044. Due to _____ 1045. Due to _____ 1046. Due to _____ 1047. Due to _____ 1048. Due to _____ 1049. Due to _____ 1050. Due to _____ 1051. Due to _____ 1052. Due to _____ 1053. Due to _____ 1054. Due to _____ 1055. Due to _____ 1056. Due to _____ 1057. Due to _____ 1058. Due to _____ 1059. Due to _____ 1060. Due to _____ 1061. Due to _____ 1062. Due to _____ 1063. Due to _____ 1064. Due to _____ 1065. Due to _____ 1066. Due to _____ 1067. Due to _____ 1068. Due to _____ 1069. Due to _____ 1070. Due to _____ 1071. Due to _____ 1072. Due to _____ 1073. Due to _____ 1074. Due to _____ 1075. Due to _____ 1076. Due to _____ 1077. Due to _____									

Death certificate coding

Disease or condition that led directly to the death

Intermediate cause of death

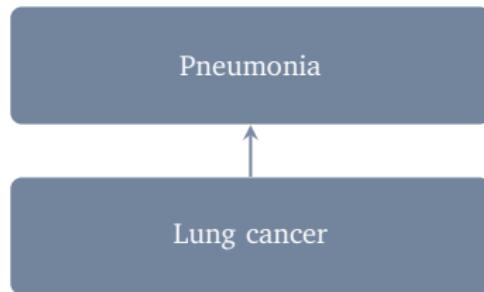
Underlying cause of death

Death certificate coding example

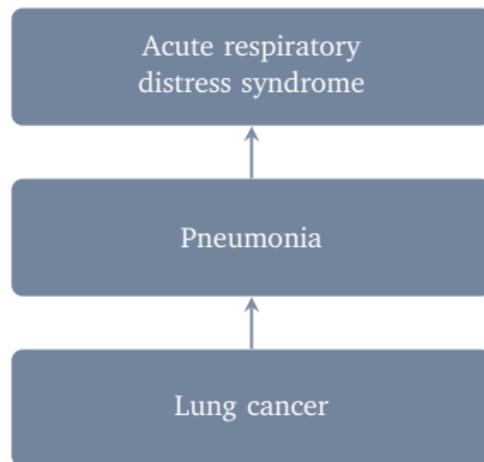
Death certificate coding example

Lung cancer

Death certificate coding example



Death certificate coding example



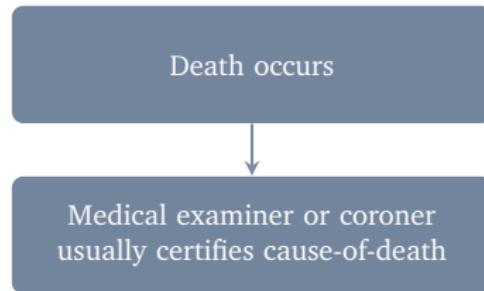
Mortality data collection process

Mortality data collection process

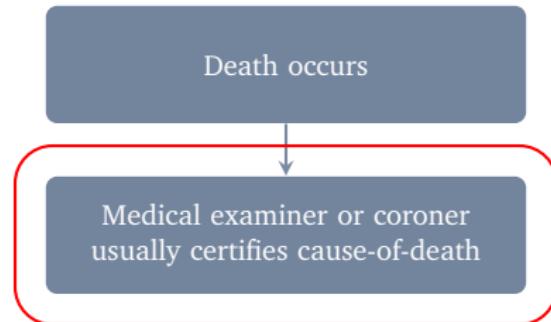
Mortality data collection process

Death occurs

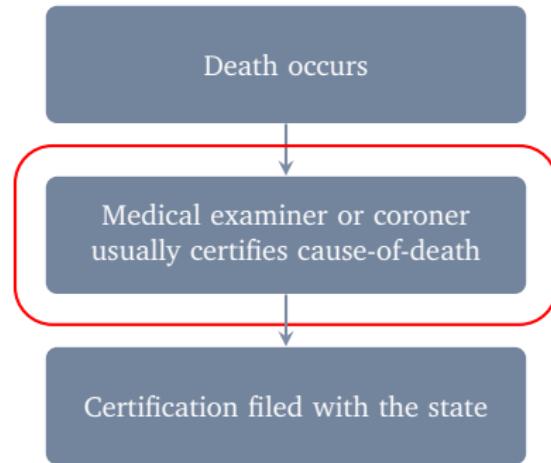
Mortality data collection process



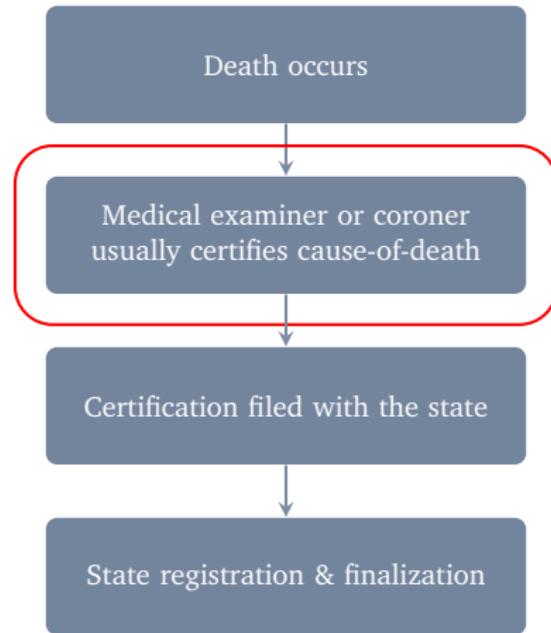
Mortality data collection process



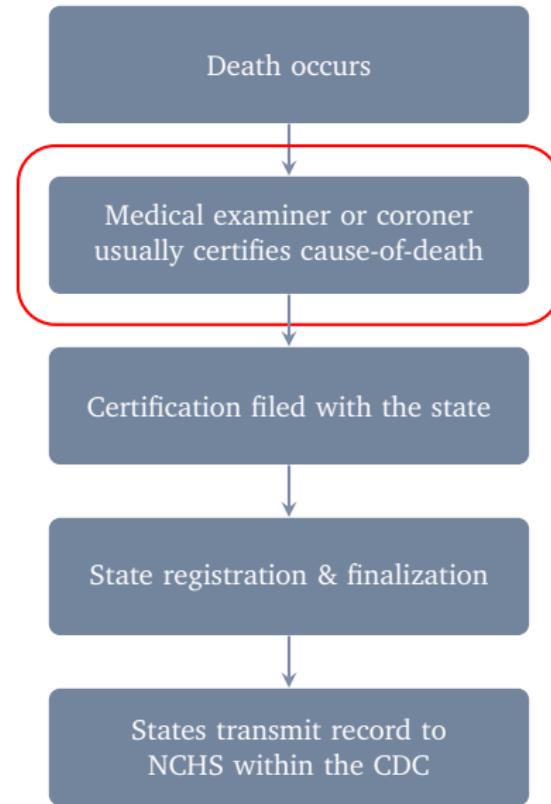
Mortality data collection process



Mortality data collection process



Mortality data collection process



Data quality problems

Data quality problems

- Missing information

Data quality problems

- Missing information
- Lack of detail

Data quality problems

- Missing information
- Lack of detail
- Garbage codes: useless or wrong codes

Data quality problems

- Missing information
- Lack of detail
- Garbage codes: useless or wrong codes

In the paper, we look at 3 aspects of data quality and then apply these metrics to the US at a county-level.

Methods

Proportion of garbage codes

Proportion of garbage codes

- First, we look at the proportion of garbage codes in each county:

$$\text{Prop. garbage} = \frac{\text{Number of garbage-coded deaths}}{\text{Number of deaths}}$$

Proportion of garbage codes

- First, we look at the proportion of garbage codes in each county:

$$\text{Prop. garbage} = \frac{\text{Number of garbage-coded deaths}}{\text{Number of deaths}}$$

- IHME has led work on garbage codes and efforts to re-classify them.

Proportion of garbage codes

- First, we look at the proportion of garbage codes in each county:

$$\text{Prop. garbage} = \frac{\text{Number of garbage-coded deaths}}{\text{Number of deaths}}$$

- IHME has led work on garbage codes and efforts to re-classify them.
- Re-assignment methods: (1) Proportional redistribution by age and sex; (2) Statistical modelling; (3) Expert judgment (for example, reconstructing the chain of events)

Proportion of garbage codes

- First, we look at the proportion of garbage codes in each county:

$$\text{Prop. garbage} = \frac{\text{Number of garbage-coded deaths}}{\text{Number of deaths}}$$

- IHME has led work on garbage codes and efforts to re-classify them.
- Re-assignment methods: (1) Proportional redistribution by age and sex; (2) Statistical modelling; (3) Expert judgment (for example, reconstructing the chain of events)
- Lozano et al, 2012 show that re-distributing garbage codes changes the top 10 leading causes of death worldwide.

Proportion of garbage codes

- First, we look at the proportion of garbage codes in each county:

$$\text{Prop. garbage} = \frac{\text{Number of garbage-coded deaths}}{\text{Number of deaths}}$$

- IHME has led work on garbage codes and efforts to re-classify them.
- Re-assignment methods: (1) Proportional redistribution by age and sex; (2) Statistical modelling; (3) Expert judgment (for example, reconstructing the chain of events)
- Lozano et al, 2012 show that re-distributing garbage codes changes the top 10 leading causes of death worldwide.
- In 2023, about 10% of deaths in the United States were garbage coded.

Level of detail

Level of detail

Goal is to measure how specific ICD-10 coding is for non-garbage codes.

Level of detail

Goal is to measure how specific ICD-10 coding is for non-garbage codes.

- We need to control for the underlying epidemiological differences in cause of death to capture the variation in detail driven by coding differences rather than the differences in the (true) distribution of causes of death. To do so, we weight deaths so that each county set matches the national average for the broad distribution of causes of death

Level of detail

Goal is to measure how specific ICD-10 coding is for non-garbage codes.

- We need to control for the underlying epidemiological differences in cause of death to capture the variation in detail driven by coding differences rather than the differences in the (true) distribution of causes of death. To do so, we weight deaths so that each county set matches the national average for the broad distribution of causes of death
- Another way to think about this is that we are interested in the diversity within each cause of death category.

Level of detail

- For each time period t , we compute the national share of deaths in each broad cause group. These national shares are used as weights.

Level of detail

- For each time period t , we compute the national share of deaths in each broad cause group. These national shares are used as weights.
- Using these weights, we construct a standardized ICD-10 distribution for each county set k :

$$p_{k,t}^*(d) = \sum_c s_t(c) w_{k,t}(d | c)$$

Level of detail

- For each time period t , we compute the national share of deaths in each broad cause group. These national shares are used as weights.
- Using these weights, we construct a standardized ICD-10 distribution for each county set k :

$$p_{k,t}^*(d) = \sum_c s_t(c) w_{k,t}(d | c)$$

- Here, $w_{k,t}(d | c)$ is the within-cause distribution of ICD-10 codes in county set k , and $s_t(c)$ is the national share of cause group c .

Level of detail

- For each time period t , we compute the national share of deaths in each broad cause group. These national shares are used as weights.
- Using these weights, we construct a standardized ICD-10 distribution for each county set k :

$$p_{k,t}^*(d) = \sum_c s_t(c) w_{k,t}(d | c)$$

- Here, $w_{k,t}(d | c)$ is the within-cause distribution of ICD-10 codes in county set k , and $s_t(c)$ is the national share of cause group c .
- $p_{k,t}^*(d)$ represents the probability that a death in county set k would be coded as ICD-10 code d , *if the county had the national cause-of-death mix*.

Level of detail

We then measure how spread out this standardized ICD-10 distribution is using Shannon entropy.

$$H_{k,t} = - \sum_d p_{k,t}^*(d) \log p_{k,t}^*(d)$$

- Entropy is a measure of diversity:
 - high entropy implies deaths spread across many ICD-10 codes
 - low entropy implies deaths clumped into a few codes

Level of detail

We then measure how spread out this standardized ICD-10 distribution is using Shannon entropy.

$$H_{k,t} = - \sum_d p_{k,t}^*(d) \log p_{k,t}^*(d)$$

- Entropy is a measure of diversity:
 - high entropy implies deaths spread across many ICD-10 codes
 - low entropy implies deaths clumped into a few codes
- We rescale entropy to a 0-100 score to obtain the level of detail.

Re-assignability Index

Re-assignability Index

The goal is to quantify how re-assignable garbage codes are in each county.

Re-assignability Index

The goal is to quantify how re-assignable garbage codes are in each county.

$$h_i = \frac{-\sum_{k \in K_g} p_i(k) \log p_i(k)}{\log |K_g|} \in [0, 1]$$

Re-assignability Index

The goal is to quantify how re-assignable garbage codes are in each county.

$$h_i = \frac{-\sum_{k \in K_g} p_i(k) \log p_i(k)}{\log |K_g|} \in [0, 1]$$

h_i := normalized Shannon entropy for record i

K_g = candidate underlying causes of death

$p_i(k)$ = probability candidate k is the true cause of death

Re-assignability Index

The goal is to quantify how re-assignable garbage codes are in each county.

$$h_i = \frac{-\sum_{k \in K_g} p_i(k) \log p_i(k)}{\log |K_g|} \in [0, 1]$$

Re-assignability Index

The goal is to quantify how re-assignable garbage codes are in each county.

$$h_i = \frac{-\sum_{k \in K_g} p_i(k) \log p_i(k)}{\log |K_g|} \in [0, 1]$$

- We calculate p_i using a model trained on the deaths with candidate i listed as the underlying cause of death and the garbage code g listed in the multiple cause of death, building on the work of Foreman et al., 2016

Re-assignability Index

The goal is to quantify how re-assignable garbage codes are in each county.

$$h_i = \frac{-\sum_{k \in K_g} p_i(k) \log p_i(k)}{\log |K_g|} \in [0, 1]$$

- We calculate p_i using a model trained on the deaths with candidate i listed as the underlying cause of death and the garbage code g listed in the multiple cause of death, building on the work of Foreman et al., 2016

$$\text{RI} := 1 - \frac{\sum_{i \in G_{c,t}} h_i}{N_{c,t}} \in [0, 1]$$

Aggregating data quality indices

Aggregating data quality indices

- Constructed an aggregate data quality index by averaging the z-scores of the three metrics for each county set

Aggregating data quality indices

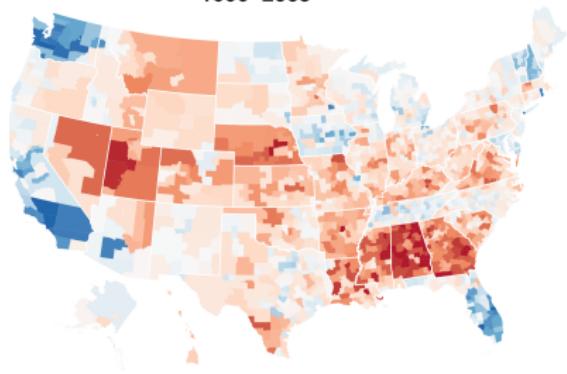
- Constructed an aggregate data quality index by averaging the z-scores of the three metrics for each county set
- For example, $z_{avg} = 1$ means the county set has, on average, one standard deviation better data quality than the mean

Results

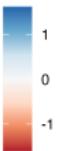
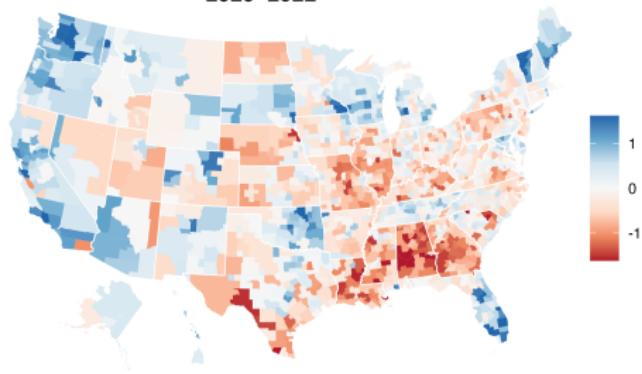
Aggregate data quality index

Aggregate data quality index

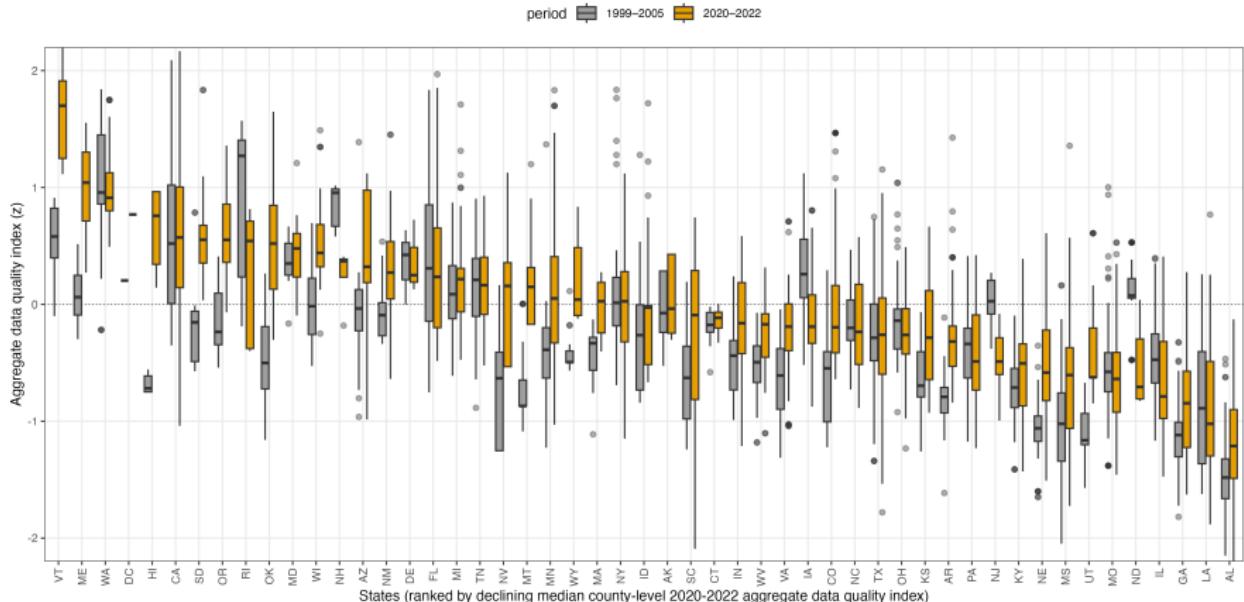
1999–2005



2020–2022



Aggregate data quality index



Aggregate data quality indices

Aggregate data quality indices

- In addition to socio-economic effects, there are clear state-specific effects

Aggregate data quality indices

- In addition to socio-economic effects, there are clear state-specific effects
- Correlation with median county income is 0.31 in 1999-2005 and 0.37 in 2020-2022

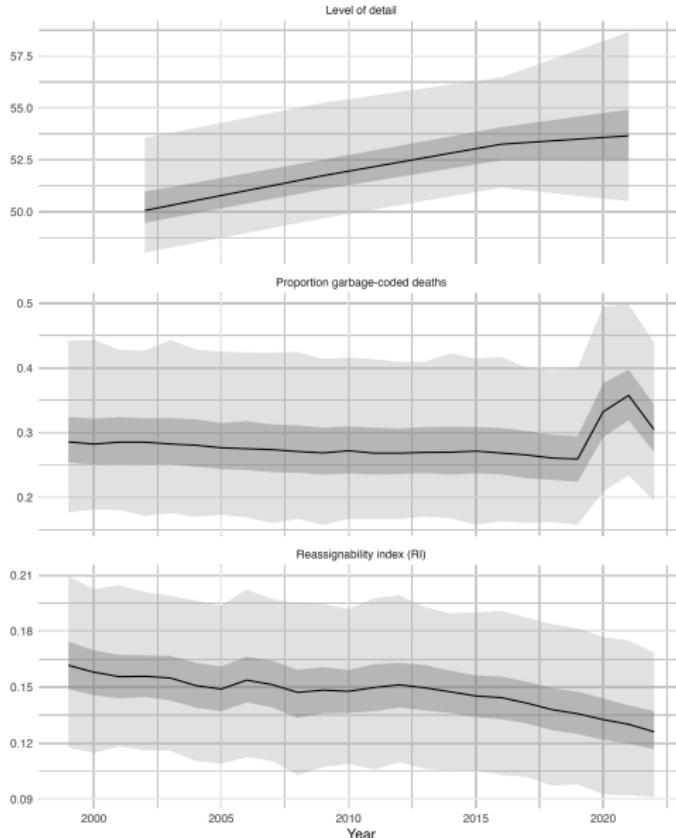
Aggregate data quality indices

- In addition to socio-economic effects, there are clear state-specific effects
- Correlation with median county income is 0.31 in 1999-2005 and 0.37 in 2020-2022
- Per-capita public health spending correlation is 0.17 in 1999-2005 and 0.14 in 2020-2022

Aggregate data quality indices

- In addition to socio-economic effects, there are clear state-specific effects
- Correlation with median county income is 0.31 in 1999-2005 and 0.37 in 2020-2022
- Per-capita public health spending correlation is 0.17 in 1999-2005 and 0.14 in 2020-2022
- Association with reporting type is 0.33 in 1999-2005 and 0.34 in 2020-2022

Data quality over time

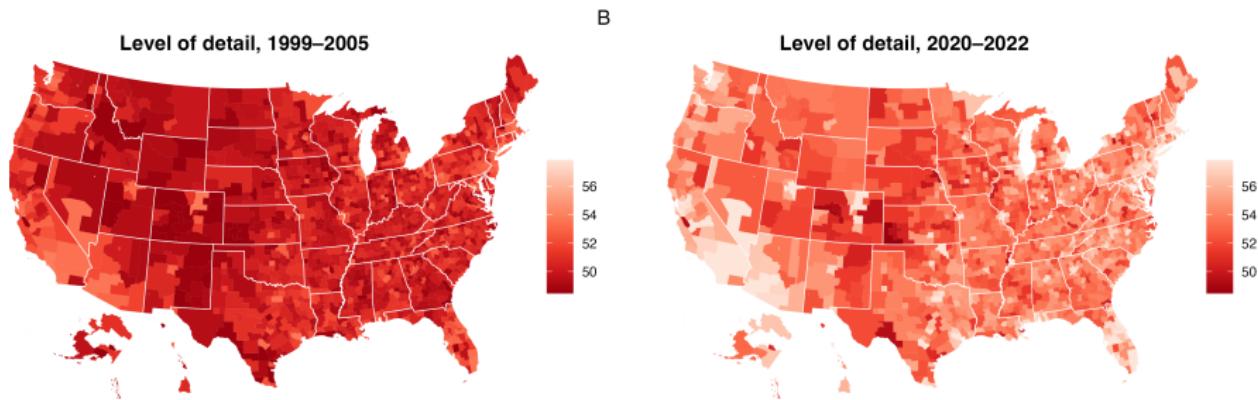


Level of detail

Level of detail



Level of detail



- Increased from 1999-2022; unclear to what extent it is driven by increasing diversity in cause of death in general (the underlying cause mixture is controlled for by period, not globally) or by improved specificity in coding

Level of detail

- Higher level of detail associated with higher median county income ($\rho = 0.23$ in 1999-2005; 0.26 in 2020-2022),

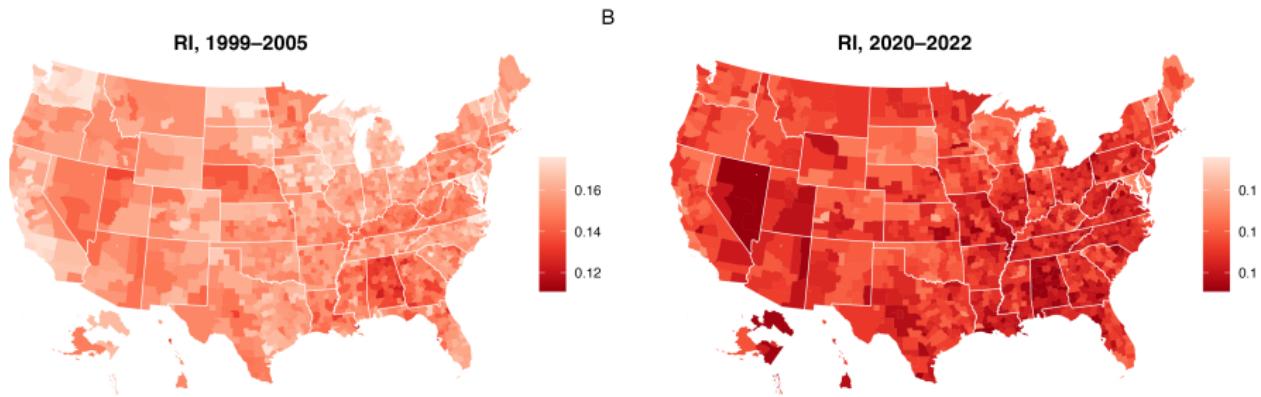
Level of detail

- Higher level of detail associated with higher median county income ($\rho = 0.23$ in 1999-2005; 0.26 in 2020-2022),
- Higher level of detail with death investigation system type ($\rho = 0.24$ in 1999-2005; 0.27 in 2020-2022).

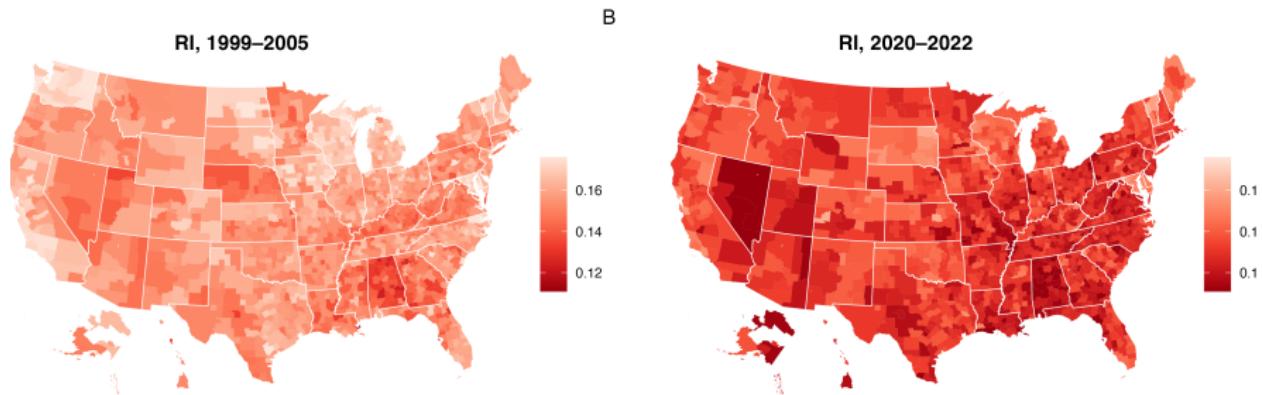
Level of detail

- Higher level of detail associated with higher median county income ($\rho = 0.23$ in 1999-2005; 0.26 in 2020-2022),
- Higher level of detail with death investigation system type ($\rho = 0.24$ in 1999-2005; 0.27 in 2020-2022).
- Small association between detail and per-capita public health spending ($\rho = 0.08$ in 1999-2005; 0.05 in 2020-2022).

Re-assignability Index

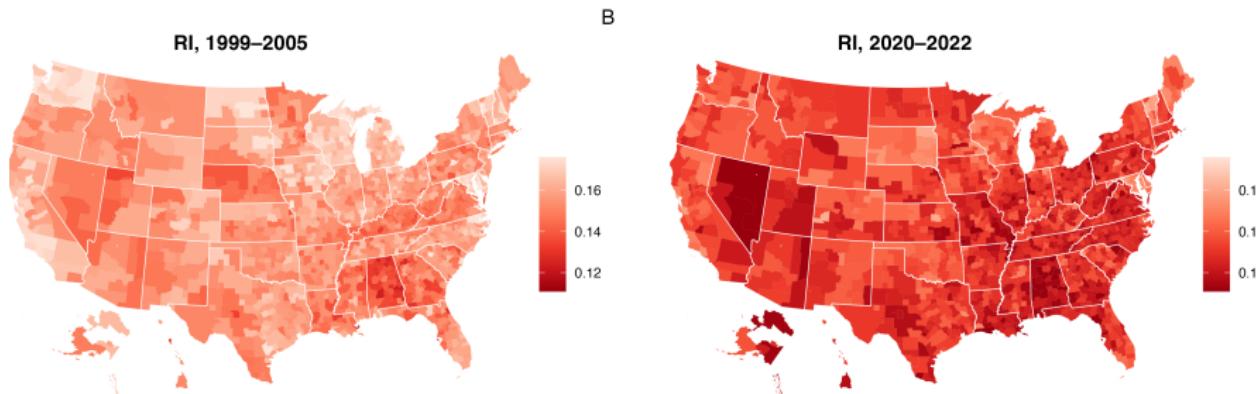


Re-assignability Index



- Positively correlated with mean county income ($\rho = 0.14$ in 1999-2005; 0.13 in 2020-2022) and with per-capita public health spending ($\rho = 0.09$ in 1999-2005; 0.08 in 2020-2022)

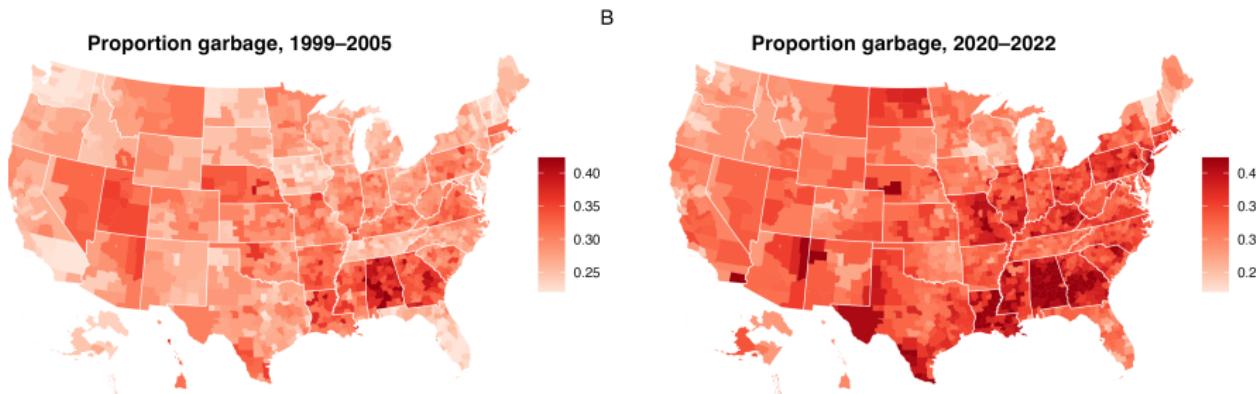
Re-assignability Index



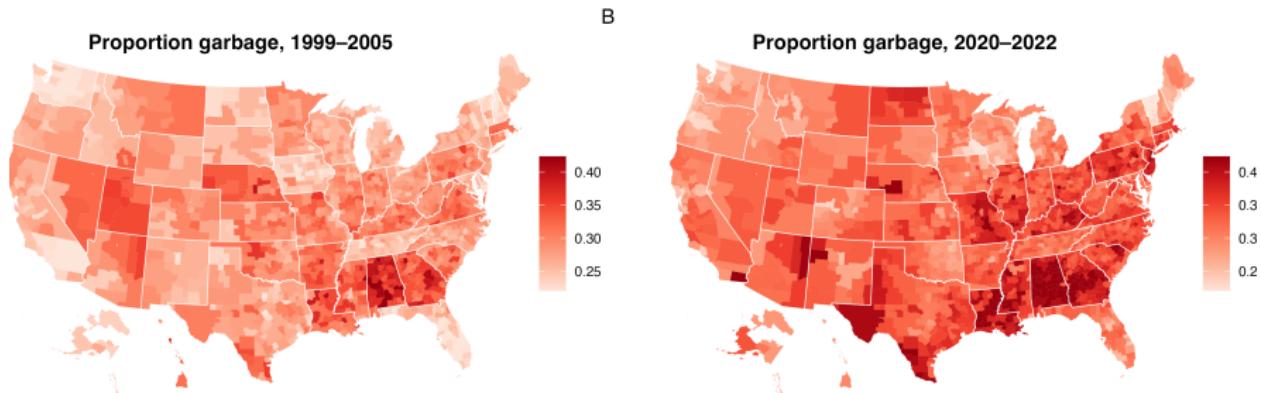
- Positively correlated with mean county income ($\rho = 0.14$ in 1999-2005; 0.13 in 2020-2022) and with per-capita public health spending ($\rho = 0.09$ in 1999-2005; 0.08 in 2020-2022)
- Small differences by reporting type ($\rho = 0.11$ in 1999-2005; 0.06 in 2020-2022).

Proportion of garbage codes

Proportion of garbage codes

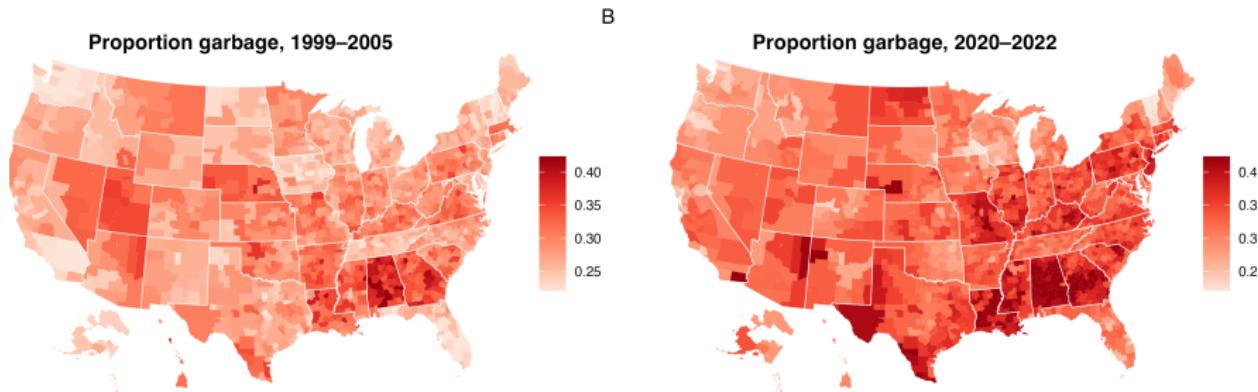


Proportion of garbage codes



- Decreased from 1999-2019 and increased during the COVID-19 pandemic
- The proportion of garbage-coded deaths was lower in higher-income counties ($\rho = -0.25$ in 1999-2005; -0.29 in 2020-2022)

Proportion of garbage codes



- Decreased from 1999-2019 and increased during the COVID-19 pandemic
- The proportion of garbage-coded deaths was lower in higher-income counties ($\rho = -0.25$ in 1999-2005; -0.29 in 2020-2022)
- Lower in counties using medical examiners ($\rho = -0.31$ in 1999-2005; -0.34 in 2020-2022) and associated with per-capita public health spending ($\rho = -0.21$ in 1999-2005; -0.15 in 2020-2022)

Relationship between data quality metrics and COVID-19 undercounting

Relationship between data quality metrics and COVID-19 undercounting

- Correlation between COVID-19 underreporting and proportion of higher garbage-coded deaths and lower level of detail is 0.27 and 0.28.

Relationship between data quality metrics and COVID-19 undercounting

- Correlation between COVID-19 underreporting and proportion of higher garbage-coded deaths and lower level of detail is 0.27 and 0.28.
- Correlation between lower RI and underreporting is 0.13

Relationship between data quality metrics and COVID-19 undercounting

- Correlation between COVID-19 underreporting and proportion of higher garbage-coded deaths and lower level of detail is 0.27 and 0.28.
- Correlation between lower RI and underreporting is 0.13
- Correlation between lower aggregate data quality index and COVID-19 underreporting is 0.33

Discussion

Discussion

- Different data quality metrics point to different potential problems

Discussion

- Different data quality metrics point to different potential problems
- More centralized reporting practises as well as use of medical examiners rather than coroners might help reduce garbage codes

Discussion

- Different data quality metrics point to different potential problems
- More centralized reporting practises as well as use of medical examiners rather than coroners might help reduce garbage codes
- Low level of detail might suggest that diagnostic specificity is constrained and could perhaps benefit from more medicolegal capacity, toxicology and autopsy access

Discussion

- Different data quality metrics point to different potential problems
- More centralized reporting practises as well as use of medical examiners rather than coroners might help reduce garbage codes
- Low level of detail might suggest that diagnostic specificity is constrained and could perhaps benefit from more medicolegal capacity, toxicology and autopsy access
- Low RI suggests incomplete or generic MCOD reporting and is likely most affected by jurisdictional differences in death certificate coding practises

Discussion

- Different data quality metrics point to different potential problems
- More centralized reporting practises as well as use of medical examiners rather than coroners might help reduce garbage codes
- Low level of detail might suggest that diagnostic specificity is constrained and could perhaps benefit from more medicolegal capacity, toxicology and autopsy access
- Low RI suggests incomplete or generic MCOD reporting and is likely most affected by jurisdictional differences in death certificate coding practises
- The strong state-specific effects on data quality is something to be optimistic about

Thank you!

Contact information:

Amy Mann

amy.mann@chch.ox.ac.uk

Monica Alexander

monica.alexander@utoronto.ca

Mathew Kiang

mkiang@stanford.edu



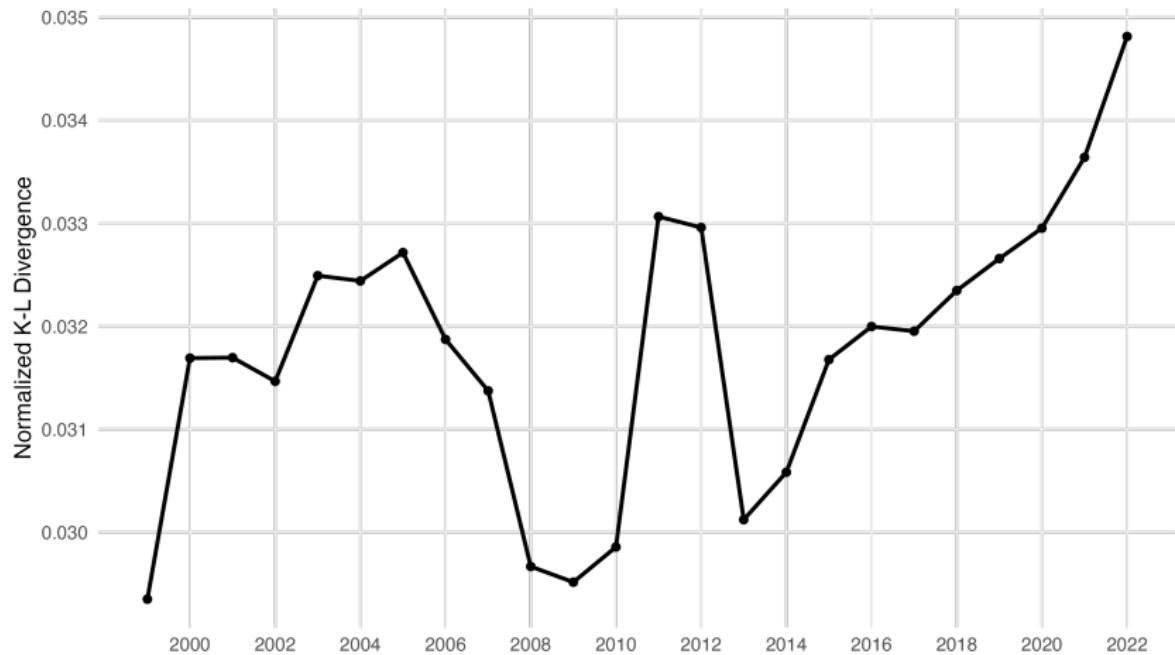
Supported by Data Sciences Institute

Supplementary Material

Why is it getting harder to re-assign garbage codes?



Why is it getting harder to re-assign garbage codes?



It's probably not because death certificates are becoming less descriptive.