Preventive Medicine 182 (2024) 107945

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Electronic nicotine delivery system use and its relation to waterpipe smoking among youth in seven Arab countries

Mohammad S. Mohammad [a], Maryam Aburezq [b], Noura AlSaeed [b], Zahraa Abdullah [c], Sarah Fayrouz [d], Khalifa Almunefi [c], Dania Bahdila [e,*]

[a] Department of General Dentistry, North Sabah Al-Salem Hospital, Ministry of Health, Kuwait
[b] Department of Pediatric, Farwanyia Hospital, Ministry of Health, Kuwait
[c] Department of Pediatric, Amiri Hospital, Ministry of Health, Kuwait
[d] Department of Pediatric, Jaber Hospital, Ministry of Health, Kuwait
[e] Department of Pediatric Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Keywords:
Electronic nicotine delivery systems
Smoking
Water Pipe Smoking
Adolescent
Arab World

ABSTRACT

Objectives: Electronic Nicotine Delivery System (ENDS) use and Waterpipe Tobacco Smoking (WTS) are reported to be a growing strain of tobacco epidemic among youth in the Arab world. Therefore, we aimed to: (1) estimate the regional prevalence of ENDS use among youth in 7 Arab countries and, (2) to explore the bidirectional relationship between ENDS use and WTS among Arab adolescents.

Methods: We analyzed data from the World Health Organization Global Youth Tobacco Survey (GYTS 2014–2018) of 18,536 schoolchildren aged 12–16 from Iraq, Mauritania, Morocco, Oman, Qatar, Tunisia, and Yemen. The weighted prevalence was calculated to generate nationally representative estimates. Adjusted multilevel logistic regression models were conducted to assess the association between ENDS use and WTS.

Results: The pooled weighted prevalence of ENDS use was 9.5%. Higher odds of ENDS use were significantly associated with WTS (AOR: 5.26, 95%CI: 4.28–6.46), smoking conventional cigarettes (AOR: 1.54, 95%CI: 1.23–1.94) and first tobacco use prior to the age of 12 (AOR: 1.40, 95%CI: 1.14–1.72). Females and children who were taught in school the dangers of tobacco had less odds of using ENDS.

Conclusion: WTS was associated with increased odds of ENDS use by >5 folds, and vice versa. Tobacco consumption at age younger than 12 years was associated with higher odds of ENDS use, but less odds of WTS. Females and those who were taught in school the dangers of tobacco were less likely to report ENDS use.

1. Introduction

Smoking tobacco products involves various forms of nicotine delivery, such as conventional cigarettes, waterpipe (narghileh, hookah, shisha. Etc.), medwakh, cigars, cigarillos, and bidis. (Abuse, 2020) Although conventional cigarettes are the most widely consumed form of tobacco smoking, sociocultural norms may influence the preference as to which tobacco method of delivery is used (Drope and Schluger, 2018). In the Arab world, waterpipe smoking (WTS) is more prevalent among women and youth than conventional cigarettes (Khalil et al., 2013). The highest level of WTS among children aged 13–15 was in Lebanon (36.9%), followed by the West Bank (32.7%) (Jawad et al., 2016). Another rising trend among youth is the use of electronic nicotine delivery systems (ENDS), specifically electronic cigarettes (e-cigarettes), which are the most used forms of ENDS (World Health Organization, 2021). ENDS are non-combusted cigarettes; they are devices that turn liquid to vapor for inhalation (World Health Organization, 2021). This liquid - known as "e-liquid" - contains tobacco and additives like nicotine or flavoring (World Health Organization, 2021).

Generally, the global prevalence of ENDS use among adolescents between 12 and 16 years is 9.8% (Sun et al., 2022). The United States remains the largest market for e-cigarettes, with 1 in 5 high school

Abbreviations: ENDS, Electronic Nicotine Delivery System; **WTS**, Waterpipe Smoking; **WHO**, World Health Organization; **GYTS**, Global Youth Tobacco Survey; **UNRWA**, United Nations Relief and Works Agency; **CC**, Conventional Cigarettes; **SHS**, Secondhand Smoke; **FCTC**, Framework Convention on Tobacco Control; **AOR**, Adjusted Odds Ratio; **95%CI**, 95% Confidence Interval; **UAE**, United Arab Emirates.

* Corresponding author: Department of Pediatric Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah 21589, Saudi Arabia.
E-mail address: dbahdila@kau.edu.sa (D. Bahdila).

https://doi.org/10.1016/j.ypmed.2024.107945
Received 8 October 2023; Received in revised form 30 March 2024; Accepted 31 March 2024
Available online 2 April 2024

students being e-cigarette users (Cornelius, 2020). Although e-cigarettes are banned in several parts of the world, including some Arab countries (World Health Organization, 2021), the outlook is changing since some countries reversed the ban allowing their import and sales (Saudi Food and Drug Authority, 2020).

Nicotine's high addiction potential poses risks to children's brain, heart, and lung health when exposed at an early age (US Department of Health and Human Services, 2016). A global meta-analysis of individuals below the age of 20 found that e-cigarette use increases the likelihood of smoking conventional cigarettes by more than twofold (Yoong et al., 2021). However, in the United States the population trends found that youth e-cigarettes use may have accelerated the decline in conventional cigarette smoking (Sokol and Feldman, 2021). The US National Academies of Sciences, Engineering, and Medicine (Eaton et al., 2018) have concluded that ENDS use is likely far less harmful than smoking combusted tobacco products. In Arab countries, concurrent consumption of various tobacco products sustains nicotine dependence and exacerbates non-communicable diseases in an already burdened region (Mokdad et al., 2014). While most epidemiological studies on e-cigarettes in the Arab world focus on adults,(Mostafa et al., 2018; Aghar et al., 2020; Hamadeh et al., 2020; Ahmed et al., 2021; Barakat et al., 2021; Karasneh et al., 2021) our study aims to estimate ENDS use prevalence among youth in 7 Arab countries and explore its bidirectional relationship with WTS. These findings will support ongoing monitoring and surveillance of the region's tobacco epidemic, informing data-driven control policies and regulations.

2. Methods

Data were obtained from the Global Youth Tobacco Survey (GYTS), a standardized multinational, self-administered, school-based survey developed to monitor tobacco smoking among adolescents. The GYTS uses a two-stage random cluster sampling method to produce a representative estimate of schoolchildren aged 13 to 15 years (Warren et al., 2000). The GYTS detailed methodology and process of ethical approval is published elsewhere (Warren et al., 2000). The full survey instruments are available at https://extranet.who.int/ncdsmicrodata/index.php/catalog/central.

2.1. Ethical consideration

The publicly available GYTS datasets have direct access provided by the World Health Organization (WHO) for researchers as part of the 4-year data turnover protocol. Ethical approval was obtained from each country's officials before conducting GYTS. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statements in our report (Von Elm et al., 2007). The study is a secondary data analysis of publicly available data. Therefore, it is considered "Not Human Subjects Research" and is exempted from the Institutional Review Board (IRB21–1156).

2.2. Inclusion and exclusion criteria

The e-cigarette module was adopted by 9 Arab countries as well as the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) in Jordan and West Bank. The inclusion criteria of a country survey were as follows: (1) the survey was the most recent survey for each country; (2) the survey was conducted on a national level to allow for cross country-level comparability; (3) the survey included data about students' e-cigarette use in the past month. We excluded UNRWA in Jordan and West Bank surveys because they were sub-national surveys. Bahrain and Kuwait were excluded because they did not ask how many days the child used e-cigarettes in the past month. We included surveys from seven countries (Iraq 2014, Yemen 2014, Morocco 2016, Oman 2016, Tunisia 2017, Mauritania 2018, and Qatar 2018). Data were extracted in 2022 and analyzed in 2023.

2.3. Measures

Current ENDS use, which was defined as using e-cigarettes at least one day in the past month, was assessed using the following question: "During the past 30 days, how many days did you use electronic cigarettes?". Current WTS, which was defined as smoking at least one day in the past month, was assessed using the following question: "During the past 30 days, on how many days did you smoke shisha".

Other predictors included current conventional cigarette (CC) smoking, which was defined as smoking a conventional cigarette at least one day in the past month, was assessed using the following question:" During the past 30 days, on how many days did you smoke cigarettes?". Age when initiated smoking: was assessed using the following question "How old were you when you first tried smoking shisha?" or "How old were you when you first tried a cigarette?". We dichotomized the variable to be "Below 12 years old" and "12 or older". We created a variable for the age when first initiated smoking regardless of whether it was CC or WTS. The student was assigned the age of whichever earlier. Secondhand smoke (SHS) exposure at home was defined as being exposed at least one day in the past week using the following questions: "During the past 7 days, on how many days has anyone smoked inside your home, in your presence?". SHS exposure in public was defined as being exposed at least one day in the past week using the following questions: "During the past 7 days, on how many days has anyone smoked in your presence, at any outdoor public place?" and/or "During the past 7 days, on how many days has anyone smoked in your presence, inside any enclosed public place, other than your home?". Learning about tobacco dangers in school was a binary variable assessed using the following question: "During the past 12 months, were you taught in any of your classes about the dangers of tobacco use?". Children who were warned about tobacco dangers in school responded by "Yes", in contrast to children who responded with "No" or "Not sure".

For other individual-level variables, we included the age (12 to 16 years old) and sex (Female or Male) of the child. For the country-level variable, we extracted the country group income (high, upper middle, lower middle, or low income) at the year of survey administration from the World Bank website. The World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) ratification year for each country was extracted from the WHO website.

2.4. Statistical analysis

To extrapolate our results to the general population, weighted estimates were reported after adjusting for the survey's weights variables: final sample weight, stratum, and primary sampling unit. Descriptive analyses were conducted to summarize the findings as sample frequencies and weighted percentages. A multilevel model was used to adjust for autocorrelation on the countries and schools' levels, where level 1 was the student, level 2 was the school, and level 3 was the country. Multilevel binary logistic models were performed to assess the association between ENDS use and WTS, after adjusting for the age, sex, conventional cigarette smoking, age when first tried tobacco smoking, SHS exposure at home, SHS exposure in public spaces, being taught in school the dangers of tobacco use and the survey administration year. The measure of association was interpreted as adjusted odds ratio (AOR) and its associated 95% confidence interval (95%CI). *P*-values <0.05 were considered statistically significant. Data were analyzed as complete case analysis using STATA version 14 (StataCorp).

3. Results

We included a total of 18,536 individuals from 216 schools, representative to 4,171,719 schoolchildren aged 12 to 16 years in 7 Arab countries, including Iraq, Mauritania, Morocco, Oman, Qatar, Tunisia, and Yemen. Data collected via GYTS were over the years 2014, 2016, 2017 and 2018. The average response rate was 89.1%, being the highest

in Tunisia (92.8%) and the lowest in Yemen (85.1%). Table 1 summarized each country's GYTS characteristic in addition to the FCTC ratification status and its income level according to the World Bank classification at the time of survey administration.

Table 2 presents the sample distribution of our sampled population. Out of 18,536 schoolchildren involved in this study aging 12 to 16 years, 5098 were 14-year-old and 4784 were 15-year-old, accounting for 29.7% and 27.8%, respectively. Males and females were almost equally involved in this study. Out of the 7 involved Arab countries, Morocco and Mauritania were the highest in the sample size, accounting for 21.1% and 20.2% respectively. Moreover, almost more than two-thirds of our study sample were of lower middle-income countries (65.9%), while the rest were of high (23.1%) and upper middle (11.0%) income level. More than half of the study sample (58.6%) started smoking at an age younger than 12 years, and approximately, 7.1% of our population were current conventional cigarette smokers. Regarding secondhand smoking, around 30.5% were exposed at home and 59.9% in public places. Additionally, 47.0% of our sample reporting being taught the dangers of tobacco use in school. The weighted prevalence of ENDS use and WTS among adolescents in each of the 7 Arab countries is illustrated in Fig. 1.

The weighted estimates of ENDS use and WTS among Arab youth is illustrated in Table 2. The weighted prevalence of ENDS use and WTS among Arab youth was 9.5% and 10.0%, respectively. Regarding current ENDS use, the highest was among those aged 15 years (2.7%), followed by those aged 14 years (2.7%). From our study sample, 6.5% of the males and 2.8% of the females were ENDS users. The highest weighted estimates of ENDS use were reported in Mauritania, Iraq and Yemen, accounting for 18.7%, 14.0% and 13.9% of each country collected sample, respectively. Schoolchildren with lower middle income resided 6.2% ENDS use, while those with high income reported 0.3% only. Among the current conventional cigarette smokers, around 2.0% used ENDS, and 8.5% of those who started smoking at an age younger than 12 years were ENDS users as well. Regarding SHS exposure, those who were exposed in public areas estimated 7.1% ENDS use, while the non-exposed ones 2.4%. On the other hand, exposure at home showed somewhat similar prevalence of WTS to those not exposed (4.2% and 5.0%, respectively).

Regarding current WTS, children aging 14 to 15 years reported the highest prevalence of WTS, 3.1% of those aging 15 years and 2.2% of the 14 years old individuals. WTS among males accounted for 7.4% and among females 2.6%. Similar to ENDS use weighted estimates, the highest WTS were reported in Mauritania, Iraq and Yemen, accounting for 18.0%, 15.3% and 14.6%, respectively. Moreover, WTS accounted for 3.0% of those smoking conventional smoking currently but 5.7% of the non-CC smokers. Of those who started smoking before the age of 12, 8.7% were WT smokers. SHS exposure in public areas was noted to have a 7.9% weighted prevalence of WTS, while non-exposed individuals reported 2.2% WTS.

Table 3 demonstrates the two multilevel binary logistic regression models of factors associated with ENDS use and WTS among Arab youth after adjustment for other confounders. In the first model, ENDS use was

Table 2
Descriptive statistics of the study population and the weighted estimates of ENDS use and WTS in school-aged children using the Global Youth Tobacco Survey (GYTS 2014–2018).

Characteristics	Sample Total N (%)	ENDS* Use Weighted Prevalence (%)	WTS* Weighted Prevalence (%)
Overall	18,536 (100.0)	9.5	10.0
Age (years)			
12	1452 (8.5)	0.4	0.4
13	3876 (22.6)	1.7	1.5
14	5098 (29.7)	2.7	2.2
15	4784 (27.8)	2.7	3.1
16	1979 (11.5)	1.4	2.0
Sex			
Male	9255 (50.4)	6.5	7.4
Female	9128 (49.6)	2.8	2.6
Current Conventional Cigarettes Smoking			
No	16,205 (92.9)	6.4	5.7
Yes	1234 (7.1)	2.1	3.0
Age when first initiated smoking (either cigarettes or waterpipe)			
Younger than 12	2631(58.6)	14.4	15.5
12 or older	1859 (41.4)	10.3	17.0
Secondhand smoke exposure at home			
No	12,604 (69.5)	5.0	4.5
Yes	5538 (30.5)	4.2	5.1
Secondhand smoke exposure in public			
No	7440 (40.1)	2.4	2.2
Yes	11,096 (59.9)	7.1	7.9
Taught at school about the dangers of tobacco use			
No	9524 (53.0)	5.2	5.3
Yes	8432 (47.0)	3.7	4.1
Survey year			
2014	4154 (22.4)	5.9	6.4
2016	6123 (33.0)	2.3	2.2
2017	2448 (13.2)	0.5	0.9
2018	5811(31.4)	0.7	0.6
Country			
Iraq	2047 (11.0)	14.0	15.3
Mauritania	3740 (20.1)	18.7	18.0
Morocco	3915 (21.1)	5.4	4.7
Oman	2208 (11.9)	6.3	9.2
Qatar	2071 (11.2)	10.8	4.4
Tunisia	2448 (13.2)	4.9	8.0
Yemen	2107 (11.4)	13.9	14.6
Living a country that ratified the World Health Organization Framework Convention on Tobacco Control			
No	3915 (21.1)	22.4	18.5
Yes	14,621 (78.9)	77.6	81.5
Country's Income according to the World Bank Classification at the year of survey administration			
Lowe Middle	12,210 (65.9)	6.2	6.4
Upper Middle	2047 (11.0)	3.0	3.3
High	4279 (23.1)	0.3	0.3

Note. ENDS: Electronic Nicotine Delivery Systems. **WTS:** Waterpipe Smoking.

Table 1
Sample Characteristics Distribution, Global Youth Tobacco Survey (GYTS 2014–2018) (n = 18,536).

Country	Sample size	Response Rate	% of boys Participants	Country's Income level*	WHO FCTC ratification status	Survey Year
Iraq	2047	88.1	66.4	UMIC	2008	2014
Mauritania	3740	91.3	45.8	LMIC	2005	2018
Morocco	3915	86.0	49.8	LMIC	–	2016
Oman	2208	91.1	48.5	HIC	2005	2016
Qatar	2071	89.0	49.9	HIC	2004	2018
Tunisia	2448	92.8	44.6	LMIC	2010	2017
Yemen	2107	85.1	53.2	LMIC	2007	2014

Note. **(*):** According to the World Bank classification at the time of GYTS administration. **WHO FCTC:** World Health Organization Framework Convention on Tobacco Control. **HIC:** High Income Country. **UMIC:** Upper Middle-Income Country. **LMIC:** Lower Middle-Income Country.

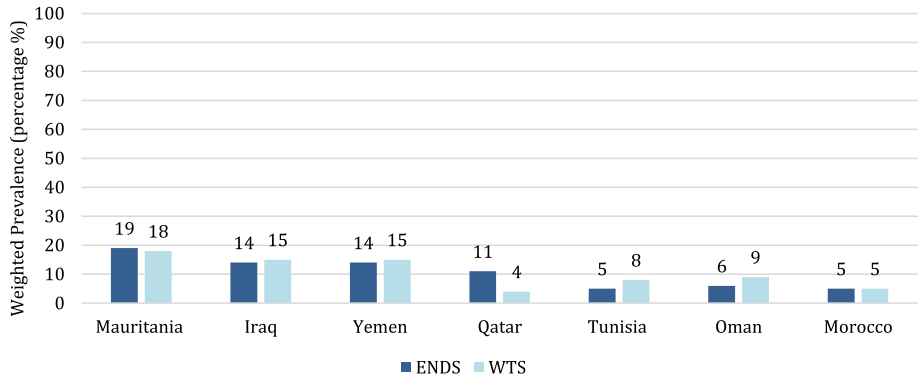


Fig. 1. Weighted prevalence of electronic nicotine delivery systems (ENDS) use and waterpipe smoking (WTS) among youth in 7 Arab countries. (Global Youth Tobacco Survey 2014–2018)

Table 3
Multilevel binary logistic regression analysis of significant factors associated with ENDS use and WTS among Arab youth (GYTS 2014–2018).

Characteristics	Model 1		Model 2	
	AOR of reporting ENDS use [95% CI]	*P* - value	AOR of reporting WTS [95% CI]	*P* – value
Current Waterpipe smoking				
No	Reference			
Yes	**5.26 [4.28–6.46]**	**< 0.001**		
Current Electronic Cigarettes Usage				
No			Reference	
Yes			**5.38 [4.36–6.63]**	**< 0.001**
Current Conventional Cigarettes Smoking				
No	Reference		Reference	
Yes	**1.54 [1.23–1.94]**	**< 0.001**	**2.63 [2.12–3.27]**	**< 0.001**
Age (years)				
12	Reference		Reference	
13	1.07 [0.67–1.72]	0.764	0.92 [0.58–1.46]	0.734
14	1.14 [0.73–1.79]	0.566	0.80 [0.51–1.24]	0.319
15	0.93 [0.59–1.47]	0.756	0.86 [0.55–1.34]	0.499
16	0.69 [0.42–1.14]	0.144	1.08 [0.67–1.75]	0.746
Sex				
Male	Reference		Reference	
Female	**0.78 [0.63–0.98]**	**0.032**	0.91 [0.73–1.14]	0.423
Secondhand smoke exposure in public				
No	Reference		Reference	
Yes	1.09 [0.86–1.39]	0.484	1.05 [0.83–1.33]	0.665
Secondhand smoke exposure at home				
No	Reference		Reference	
Yes	1.01 [0.83–1.25]	0.882	1.1 [0.91–1.35]	0.324
Age when first initiated smoking (either cigarettes or waterpipe)				
12 years old or older	Reference		Reference	
younger than 12	**1.40 [1.14–1.72]**	**0.001**	**0.78 [0.64–0.95]**	**0.015**
Taught in school the dangers of tobacco use				
No	Reference		Reference	
Yes	**0.74 [0.61–0.91]**	**0.003**	1.05 [0.87–1.28]	0.599
Survey Year	0.97 [0.81–1.17]	0.779	0.91 [0.77–1.08]	0.281

Note: GYTS: Global Youth Tobacco Survey. **ENDS:** Electronic Nicotine Delivery System. **WTS:** Waterpipe Smoking. **AOR:** Adjusted Odds Ratio. **CI:** Confidence Interval. **Bold font indicates a statistical significance.**

the outcome, and WTS was the main predictor. We found that being a current waterpipe smoker was associated with increased odds of ENDS use >5 folds compared to children who are not current waterpipe smokers (AOR 5.26, 95%CI: 4.28–6.46, $p < 0.001$). Additionally, being a CC smoker was associated with increased in the child's odds to report ENDS use compared to non-CC smoker (AOR 1.54, 95%CI: 1.23–1.94, $p < 0.001$). Children who had their first tobacco consumption before the age of 12 also have a 1.40 higher risk of reporting ENDS use compared to those started after the age of 12 years (95%CI: 1.14–1.72, $p = 0.001$). Regarding the protective factors, females were found to be significantly less likely to use ENDS in comparison to males with an odds ratio of 0.78 (95%CI: 0.63–0.98, $p = 0.032$). Moreover, receiving formal education about the harms of tobacco at school was associated with decreased odds of being ENDS users by 0.74 (95%CI: 0.61–0.91, $p = 0.003$).

In the second model, WTS was the outcome, and ENDS use was the main predictor. ENDS users were 5.38 times more likely to be WT smokers compared to non-ENDS users (95%CI: 4.36–6.63, p < 0.001). Likewise, conventional cigarette smokers were 2.63 times more likely to report WTS than non-smokers (95%CI: 2.12–3.27, p < 0.001). While those who were younger than 12 years when they first started smoking were less likely to smoke WT than those who started at older age (AOR 0.78, 95%CI: 0.64–0.95, $p = 0.015$).

4. Discussion

In this study, we aimed to estimate ENDS use prevalence among youth in 7 Arab countries and explore the bidirectional relationship between ENDS use and WTS. ENDS use prevalence among Arab adolescents was 9.5%. WTS was associated with increased odds of ENDS use by >5 folds, and vice versa. Notably, early tobacco consumption (before age 12) was associated with higher odds of ENDS use, while females and those educated about tobacco dangers in school were less likely to report ENDS use.

Regarding smoking combusted tobacco products, Reitsma et al. reported the percentage change in the prevalence of smoking among young people aged 15–24 years, from 1990 to 2019. The percentage change in smoking prevalence was as follows: Iraq (−4.98%), Morocco (−29.7%), Oman (−1.67), Tunisia (−14.0) and Mauritania (−30.5). On the other hand, it has increased in Qatar (+22.7) and Yemen (+6.95) between 1990 and 2019 (Reitsma et al., 2021).

Regarding ENDS use, the prevalence in our study mirrors the situation globally, where 9.8% of adolescents aged 12–16 reported past-30 days of use of ENDS (Sun et al., 2022). The weighted prevalence in our study was also found to vary from country to country, reaching as high as 19% in Mauritania and as low as 5% in Morocco. Regardless of the variation of ENDS use between countries, studies revealed low overall prevalence but suggested higher rates among youth and young adults (Al-Hamdani and Hopkins, 2023). A cross-sectional study conducted in

the United States reported a prevalence of 5.1% of ENDS use among the general population and 11.9% among young adults aged 18 to 24 years specifically (Boakye et al., 2022). In the Middle East, 23% of students from three universities in the United Arab Emirates self-reported current ENDS use (Abbasi et al., 2022). Moreover, 27.7% of health science students in Saudi Arabia were current ENDS users (Qanash et al., 2019), while 14% of Qatar University students were current ENDS users as well (Kurdi et al., 2021). This difference may be explained as most of the Middle East studies focused on university-based populations, unlike our population of school-aged children.

This study expands on previous reports by considering additional confounding variables affecting the odds of reporting ENDS use. Significant increase in the likelihood of ENDS use was observed among children reporting WTS, current conventional cigarette smoking, or initiating tobacco consumption before 12 years old. WT smokers reported higher ENDS use compared to conventional cigarette smokers. Although ENDS use prevalence decreased with age in our sample, the adjusted regression model revealed no statistically significant difference between age categories in the odds of ENDS use.

Additionally, we observed significantly higher odds of ENDS use among individuals not taught about tobacco dangers in school, emphasizing the protective role of anti-smoking messages in educational settings. Surprisingly, exposure to secondhand smoke (SHS) had no significant effect on raising the odds of ENDS use among schoolchildren. Similar conclusions were drawn in a study published in Tobacco Control that found no significant association between SHS exposure and ENDS use among high school students (Kong et al., 2017).

In our study, ENDS users were found to have significantly higher odds to report WTS. Also, those who smoke conventional cigarettes and those with early initiation of tobacco consumption before the age of 12 years were found to have higher odds of WTS. Previous studies in the middle east highlighted the prevalence of WTS smoking among UAE university students and how it is a current threat to young adults (Abbasi et al., 2022). Our study reaffirms the persistence of this issue in the middle east region.

Schoolchildren in low- and middle-income countries reported more ENDS use, than those in high income countries. One plausible reason could be the lack of strong tobacco control policies and regulations in these countries (Anderson et al., 2016). Low- and middle-income countries are vulnerable to advertisement and marketing of ENDS as an alternative to traditional tobacco products (Anderson et al., 2016). Moreover, ENDS are often more affordable and accessible for the lower to middle-income population compared to high-income individuals (Anderson et al., 2016). It is important to note the different perspectives on youth ENDS use among the tobacco control community. Some studies suggested that ENDS use or vaping is a method of renormalizing smoking, leading to nicotine addiction among young people (Glantz and Bareham, 2018), (Yoong et al., 2021). By contrast, other studies reported that ENDS use may assist smokers in quitting smoking (McRobbie et al., 2014) and that e-cigarettes have replaced conventional combusted cigarettes smoking among youth (Sokol and Feldman, 2021). Long-term epidemiological studies and high-quality clinical data on ENDS use health effects are relatively scarce (Balfour et al., 2021). The US National Academies of Sciences, Engineering, and Medicine (Eaton et al., 2018) have concluded that the use of non-combusted tobacco products (vaping) is likely far less harmful than combusted tobacco products.

This study has certain limitations. Firstly, we defined current smoking status as smoking at least one cigarette in the past month, which might include experimental users. However, this definition aligns with the tobacco literature, as irregular smoking among adolescents can lead to habitual smoking in the future (Kelder et al., 1994). Secondly, the study's cross-sectional design prevents us from establishing temporal or causal relationships between predictors and outcomes. Thirdly, the data included in this study were from 2014 to 2018 GYTS survey cycles. More up-to-date data are needed to reflect the evolving landscape of ENDS use among youth in the region. Additionally, not all Arab countries used the ENDS module from the GYTS, limiting the generalizability of our findings to other Arab nations.

Despite these limitations, our study utilized the validated and standardized GYTS data, allowing for a representative estimate of tobacco smoking and ENDS use among schoolchildren. Furthermore, we employed a multilevel analysis to account for clustering of data within schools and countries. This study underscores the need for ongoing monitoring and surveillance of the evolving tobacco epidemic in the region. By quantifying ENDS use and WTS prevalence and identifying associated factors, our data can inform targeted control policies on local and national levels.

5. Conclusion

The pooled prevalence of ENDS use among Arab youth from Iraq, Mauritania, Morocco, Oman, Qatar, Tunisia, and Yemen was 9.5%. WTS was associated with increased odds of ENDS use by >5 folds, and vice versa. Conventional cigarette smoking and other nicotine products consumption are reported as risk factors for both ENDS use and WTS. Tobacco consumption at age younger than 12 years was associated with higher odds of ENDS use, but less odds of WTS in the region. Females and those who were taught in school the dangers of tobacco were less likely to report ENDS use.

Authors contribution

MM, MA, NA, ZA, SF, KA, and DB designed the study, interpreted the data, drafted the manuscript, provided final approval, and agreed to be accountable for all the work in the manuscript. DB analyzed the data and is the corresponding author.

CRediT authorship contribution statement

Mohammad S. Mohammad: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Maryam Aburezq:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Noura AlSaeed:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Zahraa Abdullah:** Writing – original draft, Visualization, Methodology, Conceptualization. **Sarah Fayrouz:** Writing – original draft, Visualization, Conceptualization. **Khalifa Almunefi:** Writing – original draft, Visualization, Conceptualization. **Dania Bahdila:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors have indicated they have no potential conflicts of interest to disclose.

Data availability

Data are available at https://extranet.who.int/ncdsmicrodata/index.php/catalog/central

Must be filtered to (**Survey:** GYTS, **Countries:** Iraq, Mauritania, Morocco, Oman, Qatar, Tunisia, and Yemen).

Acknowledgments

We thank the World Health Organization and the US Centers for Disease Control and Prevention for sharing the GYTS data publicly with other researchers. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbasi, Y., Hout, M.-C.V., Faragalla, M., Itani, L., 2022. Knowledge and use of electronic cigarettes in young adults in the United Arab Emirates, particularly during the COVID-19 pandemic. Int. J. Environ. Res. Public Health 19, 7828.

Abuse, N.I.o.D, 2020. Tobacco, Nicotine, and e-Cigarettes Research Report. National Institute of Health (NIH), Bethesda, MD, USA.

Aghar, H., El-Khoury, N., Reda, M., Hamadeh, W., Krayem, H., Mansour, M., Raouf, H., Jaffa, M.A., 2020. Knowledge and attitudes towards E-cigarette use in Lebanon and their associated factors. BMC Public Health 20, 278.

Ahmed, L.A., Verlinden, M., Alobeidli, M.A., Alahbabi, R.H., AlKatheeri, R., Saddik, B., Oulhaj, A., Al-Rifai, R.H., 2021. Patterns of Tobacco Smoking and Nicotine Vaping among University Students in the United Arab Emirates: A Cross-Sectional Study. Int. J. Environ. Res. Public Health 18.

Al-Hamdani, M., Hopkins, D.B., 2023. E-cigarettes in the Middle East: The known, unknown, and what needs to be known next. Prev. Med. Rep. 31, 102089.

Anderson, C.L., Becher, H., Winkler, V., 2016. Tobacco control progress in low and middle income countries in comparison to high income countries. Int. J. Environ. Res. Public Health 13, 1039.

Balfour, D.J., Benowitz, N.L., Colby, S.M., Hatsukami, D.K., Lando, H.A., Leischow, S.J., Lerman, C., Mermelstein, R.J., Niaura, R., et al., 2021. Balancing consideration of the risks and benefits of e-cigarettes. Am. J. Public Health 111, 1661–1672.

Barakat, M., Assaf, A.M., Al-Qudah, R., Thiab, S., Alhamed, M., Al-Obaidi, H.J., Jirjees, F. J., Basheti, I., 2021. Perception of adults toward electronic cigarettes: a cross-sectional study from Jordan. Prim. Health Care Res. Dev. 22, e3.

Boakye, E., Osuji, N., Erhabor, J., Obisesan, O., Osei, A.D., Mirbolouk, M., Stokes, A.C., Dzaye, O., El Shahawy, O., et al., 2022. Assessment of patterns in e-cigarette use among adults in the US, 2017-2020. JAMA Netw. Open 5 e2223266-e66.

Cornelius, M.E., 2020. Tobacco product use among adults—United States, 2019. MMWR, Morbidity and mortality weekly report, p. 69.

Drope, J., Schluger, N.W., 2018. The Tobacco Atlas. American cancer society.

Eaton, D.L., Kwan, L.Y., Stratton, K., 2018. Public Health Consequences of e-Cigarettes.

Glantz, S.A., Bareham, D.W., 2018. E-cigarettes: use, effects on smoking, risks, and policy implications. Annu. Rev. Public Health 39, 215–235.

Hamadeh, R.R., Lee, J., Abu-Rmeileh, N.M.E., Darawad, M., Mostafa, A., Kheirallah, K. A., Yusufali, A., Thomas, J., Salama, M., et al., 2020. Gender differences in waterpipe tobacco smoking among university students in four Eastern Mediterranean countries. Tob. Induc. Dis. 18, 100.

Jawad, M., Lee, J.T., Millett, C., 2016. Waterpipe Tobacco Smoking Prevalence and Correlates in 25 Eastern Mediterranean and Eastern European Countries: Cross-Sectional Analysis of the Global Youth Tobacco Survey. Nicotine Tob. Res. 18, 395–402.

Karasneh, R., Al-Azzam, S., Nusair, M., Hawamdeh, S., 2021. Perceptions, symptoms, and practices of electronic cigarette users: Descriptive analysis and validation of Arabic short form vaping consequences questionnaire. PLoS One 16, e0245443.

Kelder, S.H., Perry, C.L., Klepp, K.-I., Lytle, L.L., 1994. Longitudinal tracking of adolescent smoking, physical activity, and food choice behaviors. Am. J. Public Health 84, 1121–1126.

Khalil, J., Afifi, R., Fouad, F.M., Hammal, F., Jarallah, Y., Mohamed, M., Nakkash, R., 2013. Women and waterpipe tobacco smoking in the eastern mediterranean region: allure or offensiveness. Women Health 53, 100–116.

Kong, G., Morean, M.E., Cavallo, D.A., Camenga, D.R., Krishnan-Sarin, S., 2017. Sources of electronic cigarette acquisition among adolescents in Connecticut. Tob. Regul. Sci. 3, 10.

Kurdi, R., Al-Jayyousi, G.F., Abdul Rahim, H.F., 2021. Prevalence, risk factors, harm perception, and attitudes toward e-cigarette use among university students in Qatar: a cross-sectional study. Front. Public Health 9, 682355.

McRobbie, H., Bullen, C., Hartmann-Boyce, J., Hajek, P., 2014. Electronic cigarettes for smoking cessation and reduction. Cochrane Database Syst. Rev. 12, Cd010216.

Mokdad, A.H., Jaber, S., Aziz, M.I.A., AlBuhairan, F., AlGhaithi, A., AlHamad, N.M., Al-Hooti, S.N., Al-Jasari, A., AlMazroa, M.A., et al., 2014. The state of health in the Arab world, 1990–2010: an analysis of the burden of diseases, injuries, and risk factors. Lancet 383, 309–320.

Mostafa, A., El Houssinie, M., Fotouh, A.A., 2018. Multiple tobacco use among young adult waterpipe smokers in Egypt. East Mediterr. Health J. 24, 7–17.

Qanash, S., Alemam, S., Mahdi, E., Softah, J., Touman, A.A., Alsulami, A., 2019. Electronic cigarette among health science students in Saudi Arabia. Ann. Thoracic Med. 14, 56–62.

Reitsma, M.B., Flor, L.S., Mullany, E.C., Gupta, V., Hay, S.I., Gakidou, E., 2021. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019. The Lancet Public Health.

Saudi Food and Drug Authority, 2020. Controls and Requirements for Electronic Smoking Services. Tobacco Control Laws. Retrieved from https://www.tobaccocontrollaws.org/files/live/Saudi%20Arabia/Saudi%20Arabia%20-%20SFDA%20E-Cig%20Requirements.pdf.

Sokol, N.A., Feldman, J.M., 2021. High school seniors who used e-cigarettes may have otherwise been cigarette smokers: evidence from monitoring the future (United States, 2009–2018). Nicotine Tob. Res. 23, 1958–1961.

Sun, J., Xi, B., Ma, C., Zhao, M., Bovet, P., 2022. Prevalence of e-cigarette use and its associated factors among Youths aged 12 to 16 years in 68 countries and territories: global youth tobacco survey, 2012–2019. Am. J. Public Health 112, 650–661.

US Department of Health Human Services, 2016. E-cigarette use among youth: A report of the surgeon general—Executive summary. Off. Smok.

Von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bull. World Health Organ. 85, 867–872.

Warren, C.W., Riley, L., Asma, S., Eriksen, M.P., Green, L., Blanton, C., Loo, C., Batchelor, S., Yach, D., 2000. Tobacco use by youth: a surveillance report from the Global Youth Tobacco Survey project. Bull. World Health Organ. 78, 868–876.

World Health Organization, 2021. WHO report on the Global Tobacco Epidemic 2021: Addressing New and Emerging Products.

Yoong, S.L., Hall, A., Turon, H., Stockings, E., Leonard, A., Grady, A., Tzelepis, F., Wiggers, J., Gouda, H., et al., 2021. Association between electronic nicotine delivery systems and electronic non-nicotine delivery systems with initiation of tobacco use in individuals aged< 20 years. A systematic review and meta-analysis. PLoS One 16, e0256044.

ANNUAL REVIEWS

Annual Review of Public Health

Innovations in Public Health Surveillance for Emerging Infections

Peng Jia,[1,2,3,4,*] Shiyong Liu,[5,*] and Shujuan Yang[6,4]

[1]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China; email: jiapengff@hotmail.com

[2]Hubei Luojia Laboratory, Wuhan, China

[3]School of Public Health, Wuhan University, Wuhan, China

[4]International Institute of Spatial Lifecourse Health (ISLE), Wuhan University, Wuhan, China

[5]Center for Governance Studies, Beijing Normal University at Zhuhai, Zhuhai, China

[6]West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China; email: rekiny@126.com

Annu. Rev. Public Health 2023. 44:55–74

First published as a Review in Advance on January 10, 2023

The *Annual Review of Public Health* is online at publhealth.annualreviews.org

https://doi.org/10.1146/annurev-publhealth-051920-093141

*These authors contributed equally to this article

Keywords

public health surveillance, national notifiable disease reporting system, infectious disease, COVID-19, spatial modeling, artificial intelligence

Abstract

Public health surveillance is defined as the ongoing, systematic collection, analysis, and interpretation of health data and is closely integrated with the timely dissemination of information that the public needs to know and upon which the public should act. Public health surveillance is central to modern public health practice by contributing data and information usually through a national notifiable disease reporting system (NNDRS). Although early identification and prediction of future disease trends may be technically feasible, more work is needed to improve accuracy so that policy makers can use these predictions to guide prevention and control efforts. In this article, we review the advantages and limitations of the current NNDRS in most countries, discuss some lessons learned about prevention and control from the first wave of COVID-19, and describe some technological innovations in public health surveillance, including geographic information systems (GIS), spatial modeling, artificial intelligence, information technology, data science, and the digital twin method. We conclude that the technology-driven innovative public health surveillance systems are expected to further improve the timeliness, completeness, and accuracy of case reporting during outbreaks and also enhance feedback and transparency, whereby all stakeholders should receive actionable information on control and be able to limit disease risk earlier than ever before.

INTRODUCTION

Public health surveillance is defined as the ongoing, systematic collection, analysis, and interpretation of health data that are essential to the planning, implementation, and evaluation of public health practice and is closely integrated with the timely dissemination of information that the public needs to know and upon which the public should act (75). Public health surveillance is central to modern public health practice by contributing data and information usually through a national notifiable disease reporting system (NNDRS), which, although named differently in various countries (e.g., **Table 1**), often refers to the case surveillance system for infectious and other reportable conditions (23). Among all functions of public health surveillance, early identification and accurate forecasting of the timing, intensity, and distribution of emerging infectious diseases have been of high priority. This is particularly true in the current context of the coronavirus disease 2019 (COVID-19) pandemic, which was caused by a novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 was declared as a public health emergency of international concern on January 30, 2020, and further as a pandemic by the World Health Organization (WHO) on March 11, 2020.

The numbers of identified COVID-19 infections and deaths, distributed in 213 countries and territories, reached more than 538 million and 6.3 million by June 22, 2022, respectively. As of the publication date of this article numbers are still going up worldwide, even in those countries best prepared to deal with a pandemic, according to the 2019 Global Health Security Index, which measures health security and related capabilities with regard to epidemic prevention, early detection for epidemics, rapid mitigation of epidemic spread, the presence of a robust health system to treat the sick, compliance with international norms, and country vulnerability to epidemics (10). In addition, although the pandemic was generally contained in China during March–April 2020, localized outbreaks of COVID-19 have emerged from time to time since June 2020, some of which were due to the influx of contaminated meat and seafood from abroad (e.g., in early July 2020, Chinese customs found traces of SARS-CoV-2 on shrimp packaging from Ecuador at Dalian and Xiamen ports). Such incidents have raised concerns about imports of frozen products and the inadequacy of monitoring only traveler flows (94). Moreover, the emerging SARS-CoV-2 variants occurring first in the United Kingdom and India have also been exported to many other countries via air travel. All these lessons learned from COVID-19 imply that the current global networks of NNDRSs need to be innovated in all countries to better protect one another in the inseparably intertwined larger international network.

Although early identification and prediction of future disease trends may be technically feasible, more work is needed to improve accuracy so that policy makers can use these predictions to guide prevention and control efforts, as described in a recent call motivated by COVID-19 for China to upgrade its current NNDRS (34). This article consists of four sections. In the first two sections, we review the advantages and limitations of the current NNDRS in most, if not all, countries. In the third section, we discuss some lessons learned about prevention and control from the first wave of COVID-19. In the last section, we describe some technological innovations in public health surveillance, including information technology, data science, and the digital twin method.

MAJOR ADVANTAGES OF THE NNDRS

In most countries, the NNDRS is a passive surveillance system that collects information on infectious diseases and usually has a built-in geographic information system (GIS) to visualize spatial information of reported disease cases and plot their clusters to support areal-level analysis for infectious disease outbreaks. Countries have been moving toward electronic reporting through the NNDRS, which has improved timeliness. Although the current global networks of NNDRSs have

Table 1 Advantages and limitations of the national notifiable disease reporting systems (NNDRS) in the ten countries best prepared to deal with a pandemic, according to the 2019 Global Health Security Index (from high to low)

Countries	NNDRS	Advantages	Limitations
United States	NBS, NNDSS, NORS	NBS: ■ Reduced time of reporting diseases ■ Receiving more laboratory reports ■ Improved communication among local, state, and federal public health staff ■ Ability to push data entry back to the sources ■ Reduced paper-based reporting ■ Robust reporting modules	NNDSS: ■ Differential management of the NNDSS across states ■ Lack of epidemiological data exchange across different levels of the CDC without building extra infrastructure ■ Poor communication and coordination among different levels of the CDC ■ Poor communication and coordination among the electronic medical records and case reporting systems of different levels of public health sectors NORS: ■ Voluntary reporting standards for diseases ■ A small proportion of cases related to outbreaks ■ Unknown ability to reflect the same sources of infection and settings of sporadic diseases ■ Unreported outbreaks on cruise ships at the international ports and outside the United States
United Kingdom	NOIDS	■ Electronic notification of diseases with an electronic signature for authorization of general practitioners ■ Reduced delay through computerization of storage and retrieval of data	■ Lack of epidemiological information at a regional level ■ Many manually performed functions
The Netherlands	Osiris	■ Emphasizing the concept of one health ■ Inclusion of the first Center for One Health worldwide ■ Focusing on syndromic surveillance	■ A risk of overlooking an outbreak ■ Missing complete notification data
Australia	NNDSS	■ Reduced time of reporting diseases to jurisdictional notification (from fortnightly to daily) ■ Reduced time of data upload from jurisdictions to NNDSS (from fortnightly to daily)	■ Lack of epidemiological data exchange across different levels of the CDC without building extra infrastructure ■ Poor communication and coordination among different levels of the CDC ■ Fragmentation of data collection and incompatible notifiable disease databases

(*Continued*)

Table 1 (*Continued*)

Countries	NNDRS	Advantages	Limitations
Canada	CNDSS, GPHIN	CNDSS: ■ Improved interactivity of the Web-based system ■ Provides open data on all 56 notifiable diseases GPHIN: ■ A high-performing distributed system that can afford the big data ■ Integration of multisource public health information ■ Timely reporting of infectious diseases	CNDSS: ■ Voluntary reporting standards for diseases GPHIN: ■ Limited scalability due to the unavailable knowledge sources ■ Unverified and differential data compared with the true reported cases ■ Limited data sources from media in English or French only ■ Inaccurate data from media in other languages
Thailand	NADSS, National Surveillance System	NADSS: ■ Annual sero-surveillance in dairy cattle ■ Enhanced veterinary capability in the early warning system ■ Compliance with the WOAH international animal health codes ■ Screening for transboundary animals	National Surveillance System: ■ Voluntary reporting standards for diseases ■ Varied reporting methods and requirements over time
Sweden	SmiNet	■ Easy data collection based on the Internet-based forms ■ Readily accessible, cost-effective, and scalable ■ Timely data flows ■ Full integration of clinical and laboratory notification ■ High performance in handling more than 50 diseases	■ Not being able to cover the entire population ■ No laboratory testing included ■ Need for recruiting and maintaining participants
Denmark	Danish surveillance registry	■ High data quality ■ General population surveillance	■ Extensive efforts in comparing and validating the same information/data from multiple registries ■ Completeness of the diagnosis (relative to the general population) depends on whether the condition requires hospitalization and on diagnostic and coding practices ■ Less supervised registry enrollment and bias for registries due to missing or incomplete data ■ Extensive and constant financial support needed for implementation, operation, and maintenance of the system

(*Continued*)

Table 1 (*Continued*)

Countries	NNDRS	Advantages	Limitations
South Korea	NNDSS, Disease Web Statistics System	NNDSS: ■ Timely reporting The Disease Web Statistics System: ■ Standardized and informative reporting of the national notifiable diseases, including diagnostics and epidemiological information of each infectious disease ■ Data sharing with health care agencies	NNDSS: ■ Time lag in diagnosis ■ Delayed reporting by doctors ■ Weak public education and clinical guidelines The Disease Web Statistics System: ■ Voluntary reporting standards for diseases
Finland	FNIDR	■ Whole population surveillance ■ Saving manpower in the laboratory-based notification through automated computer algorithms	■ Inconsistencies in the information systems of different service providers ■ High challenge in data sharing due to lack of integration across different information systems ■ Data updates on an annual or monthly (rather than weekly or daily) basis

Abbreviations: CDC, Centers for Disease Control and Prevention; CNDSS, Canadian Notifiable Disease Surveillance System; FNIDR, Finnish National Infectious Diseases Register; GPHIN, Global Public Health Intelligence Network; NADSS, National Animal Disease Surveillance System; NBS, National Electronic Disease Surveillance System Base System; NNDRS, National Notifiable Diseases Reporting System; NNDSS, National Notifiable Diseases Surveillance System; NOIDS, Notifications of Infectious Diseases; NORS, National Outbreak Reporting System; WOAH, World Organization for Animal Health.

not performed adequately in preventing and controlling COVID-19, many of these systems have important advantages (**Table 1**).

The NNDRS has at least three advantages compared with the early paper-based disease reporting systems. First, data collection of infectious disease cases has been simplified. For example, prior to the NNDRS, district/county-level disease cases were aggregated and reported by mail only to the upper-level Center(s) for Disease Control and Prevention (CDC) in China (i.e., the municipal/prefecture CDC) and then to the provincial CDC and finally to the national CDC. The NNDRS changed this operational mechanism in 2004 by enabling electronic reporting of individual cases from all hospitals and primary health care clinics directly to the national CDC. Second, the timeliness of reporting has been substantially improved, which could assist in forecasting disease outbreaks and designing control strategies. For example, the National Electronic Disease Surveillance System Base System (NBS), developed by the US CDC in 2002, has collected more than 700,000 notifiable disease cases and provided timely information for the US CDC to use in making decisions in response to public health emergencies (12). This advantage has also been observed in other countries, including the NNDRS in China (33), the National Notifiable Disease Surveillance System (NNDSS) in Australia (3), the NNDSS in South Korea (91), and the National Epidemiological Surveillance of Infectious Diseases in Japan (58), where the times of updates, data analyses, and communication on diagnostics of the reported cases have been reduced from time periods of, e.g., 1–2 weeks, to a daily basis. Third, disease surveillance has been more cost-effective as case collection and reporting have been digitalized and the surveillance of multiple infectious diseases can be managed in one system (12).

The data quality and reliability of the NNDRS have been improved by distributed computing technology, high standards for constructing disease databases, and the scalability of disease surveillance systems. For instance, the Danish surveillance registry is of high data quality and is

often used in connection with research projects because the system can automatically check data completeness and accuracy from multiple registries (65, 74, 77); with systematic and hierarchical reporting and disseminating mechanisms, the national integrated surveillance system in Italy enables investigators to predict the burden of disease and incidence trends for each hepatitis type on the basis of complete case data, including serological testing results, sociodemographic characteristics, geographic location, incidence rate, and information on risk factors (78). In addition, the inclusion of a "one health" component, focusing on the human–animal–environment disease interface (30), in the NNDRS has improved the robustness of the NNDRS in facing emerging infectious diseases and has strengthened the ability to investigate the emerging sources of infectious diseases. About 60.3% of the emerging infectious diseases reported during 1940–2004 belonged to zoonotic diseases, 71.3% of which, including COVID-19, originated in wildlife (36). Hence, the one health approach is extremely important for the early warning of future epidemics (10), which has been initially realized in some countries. For example, the Netherlands Center for One Health was established to report the potential zoonotic threat monthly to the Dutch National Institute for Public Health and the Environment Center for Infectious Disease Control (20, 80).

MAIN LIMITATIONS OF THE NNDRS

Despite the aforementioned advantages, several limitations to NNDRSs across the world contributed to the heavy loss of lives due to COVID-19 globally. First, NNDRSs, on average, are insufficiently timely. The current NNDRSs in most countries have at best a built-in, perhaps sometimes Web-based, GIS for real-time visualization of spatiotemporal dimensions of diseases and epidemics, e.g., temporal trends and/or geographical distribution of disease cases. This approach, however, is still considered retrospective due to the inherent time lag in case reporting and thus is insufficient to presciently and robustly identify early risk and issue early warnings for public health emergencies. Moreover, as of 2019, only 32% of all countries had an interoperable electronic real-time communicable disease surveillance system, and only about 11% of African countries were considered to have acceptable real-time surveillance and reporting systems (10); the WHO emphasized the need for preparedness for COVID-19 in African countries at the beginning of the pandemic due to their vulnerable public health surveillance systems. This problem and the subsequent reporting issues have partly accounted for the underestimation of epidemics by governments in both developed and developing countries (in either participant-based or health care–based reporting systems), such as the National Outbreak Reporting System (NORS) in the United States (**https://www.cdc.gov/nors/data/using-nors.html**), the NNDSS in Canada (73), the National Electronic Surveillance System (SmiNet-1) in Sweden (85), the National Surveillance System in Thailand (48), and the Disease Web Statistics System in South Korea (59).

Second, communication and coordination among different levels of the CDCs within countries are insufficient. Race/ethnicity concerns (15), privacy (53), and ethical issues (39) are primary challenges to inhibiting data sharing across surveillance systems at different levels. This lack of data sharing has hindered our ability to understand the epidemic trend, which usually varies at different geographic scales and across countries worldwide. Technical barriers, manifested in the incompatibility and standardization issues for data coming from different electronic health record (EHR) systems and the lack of process interoperability in health care systems, have also hindered effective data sharing across clinical and public health fields (1, 26, 42). For example, the NNDRS could not exchange epidemiological data across CDCs at different levels without building extra infrastructure for data exchange, such as in the United States, Australia, Germany (3), and the United Kingdom (13). The direct influences of these phenomena included the fragmentation of data collection and the incompatible notifiable disease databases, which could hinder accurate monitoring of epidemiological data on epidemics. This problem has also been highlighted in

the 2014 Australian National Framework for Communicable Disease Control (2). In the United States, multijurisdictional and multihierarchical data sharing across different states has been facing legal barriers because state disease reporting laws prevent the sharing of personally identifiable health information across jurisdictions (37). The United States has three communicable disease surveillance systems: NORS, NBS, and NNDSS. However, it is difficult to obtain uniform epidemiological data from the US CDC; this was especially true during COVID-19. Therefore, data-sharing protocols or agreements between states and countries may also create barriers to fully utilizing state-level data in surveillance systems at different hierarchies (15). In addition, data sharing across countries faces even more technical, motivational, economic, political, legal, and ethical challenges than data sharing across jurisdictions in one country (49, 79).

Third, all existing early-warning systems require ongoing monitoring and evaluation. This need was made apparent when these systems failed to identify the COVID-19 outbreak at the early phase. Although countries including the Netherlands and China have taken steps toward early notification of emerging infectious disease outbreaks [e.g., the Dutch legislation and the Chinese pneumonia surveillance system (80)], the ongoing lack of clarity regarding standards for notification and hence the underreporting of issues limit their capacity to prevent emerging infectious disease outbreaks. Also, although the Dutch early-warning system, which relies on expert opinions, has functioned well because the number of outbreaks has been limited and communication lines between governments and medical professionals are short, there is always a risk of overlooking disease outbreaks when the number of outbreaks increases, such as with COVID-19 (80). Syndromic surveillance has been proposed as an investigational approach that uses symptom and/or preliminary diagnosis information and rapid data collection methods to provide information for public health action in a more timely manner than other more traditional approaches. However, several challenges prevent its ability to function fully, such as difficulties in defining optimal data sources, evaluating appropriate syndromic definitions, and developing minimally acceptable response protocols. Given the intrinsic trade-offs among timeliness, sensitivity, and false alarm rate (72), as well as concerns for privacy and possible public panic, relevant agencies and governments remain conservative in deploying personnel and financial resources to implement syndromic surveillance and hesitate to activate it if it exists (5, 72). Another challenge is differentiating emerging infectious diseases from other known diseases with similar symptoms, when using generic respiratory syndromic indicators, such as those used to distinguish COVID-19 from influenza (18). Moreover, syndromic surveillance cannot identify asymptomatic cases.

Fourth, all existing NNDRSs lack a strong module that focuses on the linkage among human, animal, and environmental data. For instance, data on human, animal, and wildlife surveillance can be shared across different ministries in only 30% of all countries, with a lack of data-sharing mechanisms in other countries (10). Interoperability, convergent integration, semantic consistency, and interconnectivity are four primary mechanisms of integration between human and animal health surveillance systems (21).

Fifth, epidemiological data are still disconnected from real-time laboratory data. About 77% of all countries lack the ability to collect ongoing or real-time laboratory data (10). In the other 23% of countries, this procedure should be expedited by modern technologies to prevent the spread of highly transmissible diseases, such as COVID-19, which has advanced the rapid synchronization of NNDRSs and laboratory data (83). However, due to the requirements for technological capability (genome sequencing for mutant tracing), the significant financial investment needed to acquire necessary equipment (e.g., RT-PCT machine), and the advanced knowledge and personnel needed for processing and integrating epidemiological and laboratory data, the real-time connection of epidemiological data and laboratory data remains very challenging even in developed countries, not to mention in less developed countries (45, 57).

Finally, the capacity for small area estimation (SAE), i.e., estimating parameters in small geographical areas or in small subpopulations of interest (with few or no available samples) included in a larger survey, is still limited in many NNDRSs. This lack of capacity limits the opportunity to quickly investigate disease patterns in a small geographic area or subpopulation. A small-area surveillance module using the SAE approach should be integrated with the NNDRS, which can easily collect data (e.g., case reports) and conduct modeling (e.g., Bayesian spatial models) at any small geographic scale, carry out the accuracy assessment of SAE, and disseminate SAE results to policy makers and the public (81, 95). Such small-area surveillance modules enable public health agencies to identify spatiotemporal patterns of infectious diseases at a local geographic level, investigate the underlying reasons, and implement precision containment strategies accordingly (e.g., defining the extent of lockdown and supply of vaccines) (43).

LESSONS ABOUT PREVENTION AND CONTROL FROM THE FIRST WAVE OF COVID-19

Early Case Identification

Early case identification is key to controlling and mitigating an epidemic. Detection capacity is largely a function of well-funded public health surveillance programs integrated with robust health care systems. Recent technological developments, such as EHRs, GIS, and the analytic capacity for real-time monitoring, can play key roles in quickly identifying new cases and deploying appropriate responses (e.g., contact tracing). Vertical integration of the public health capacity, from country level to community and hospital levels, is also critical as successful early case identification relies on not only local communities but also the data-linkage capacity from local to regional and national surveillance. For example, in the United States, hospital systems often have rapid high-quality data; however, these do not connect well to the public health surveillance infrastructure. Although the United Kingdom has extensively used EHRs to help support COVID-19 surveillance and containment, fewer efforts have been made to establish centralized data collection, integrate validation mechanisms across the linked EHRs, and implement rapid synchronization mechanisms with NNDRSs because this work is time-consuming and resource-intensive (86). Lacking such efforts may have caused selection bias and over- or under-interpretation of relevant findings and therefore significantly lowered the value of EHRs in informing NNDRSs (9).

Testing and Tracing Capacity

In addition to early detection, countries that were able to rapidly deploy intense testing capacities fared better during the COVID-19 pandemic (89). Many countries experienced significant delays in their expansion of testing capabilities for various reasons, including poorly planned infrastructure, disruptions in supply chains (e.g., for testing reagents), and slow policy response. The implementation of contact tracing, which is typically effective yet highly labor-intensive, has varied considerably by country. In addition, many countries or jurisdictions may have had the capacity to collect testing and tracing data at large scales and even isolate close contacts, but they did not have the capacity to analyze the epidemiological patterns of testing and tracing data.

Capacity to Implement Required Public Health Policies

During the COVID-19 pandemic, a given country's response capacity has been strongly linked to the roles, responsibilities, and structural components of its public health agencies. Overall, countries with a history of strong public health investments and experience in managing interventions successfully reacted not only more quickly, but also more effectively and were better

prepared to enact traditional outbreak response strategies, including isolation, quarantine, social distancing, and community containment (84, 88). Networks and international collaboration have been improving, with lessons learned from previous epidemics such as severe acute respiratory syndrome (SARS) and with structures such as the Global Outbreak Alert and Response Network (GOARN). Data-sharing consortiums, such as GISAID (54), have also been playing an increasingly important role in the rapid sharing of information and strategies.

Country-Wide Practice and Policy Response

The impacts of COVID-19 have highlighted the importance of resilience planning, which could ensure that systemic impacts from pandemics and other natural or human-made disasters are minimized in terms of scope and severity, especially with regard to vulnerable populations. The observed differences in COVID-19 impacts among countries have highlighted the shortcomings of respective country-level institutions in terms of public health surveillance, organization, and health care capacity and more broadly across economic, social, and environmental factors. Although international comparisons are not straightforward, there have been marked variations in government readiness and the ability or willingness to take decisive and comprehensive action to stave off the worst effects of the pandemic. Germany and South Korea performed relatively well, quickly expanding their testing and tracing capabilities. By contrast, the United Kingdom, the United States, and Brazil, among others, were slow to react and haphazard in the policies they adopted. This approach was reflected in delays in enacting lockdown procedures, severe shortfalls in testing and tracing, low stocks of personal protective equipment, and confusion in terms of public health communication. These serious deficiencies in pandemic planning and response sparked protests and have often been accompanied by poor transparency and resistance to accountability (for instance, no major policy makers in any of the three countries mentioned above have been removed from their positions) (41).

Systems Thinking to Drive Holistic Public Health Surveillance

While health emergency response has relied primarily on the health sector (e.g., public health surveillance, public health agencies, and health care organizations), some epidemic containment policies, such as lockdowns during the COVID-19 pandemic, have had major and ongoing social, economic, environmental, and behavioral impacts (29, 35, 51, 92, 93). Systems thinking is required to gain a holistic understanding of those impacts because of the complex links and interactions among multiple domains by arranging a set of interacting and interdependent elements that function as a whole, producing what individual constituents cannot produce (50, 52, 61). Systems thinking offers a paradigm of understanding health as a structured system exhibiting dynamic complexity over time within certain contexts. In the design and operation of public health surveillance, systems thinking can (*a*) help stakeholders understand how disease spreads as people interact with each other and their contexts within social networks; (*b*) encourage people to transcend disciplines and organizations to pursue understanding of more complicated health system challenges from multiple angles; (*c*) conceptualize structure design of new surveillance systems and strategize major changes to better cope with public health challenges; (*d*) guide the design of infrastructure to promote seamless connections among EHR systems, national-level lab databases, and surveillance systems at national and local levels; and (*e*) facilitate coordination and communication among lab, clinical, hospital, surveillance, public advising, and public health administration teams (17, 44). Such approaches can provide leverage points for interventions and have shown potential for addressing inequalities (27, 64). The COVID-19 pandemic has highlighted the need for reforms and improvements to the capacity of both global and national public health governance,

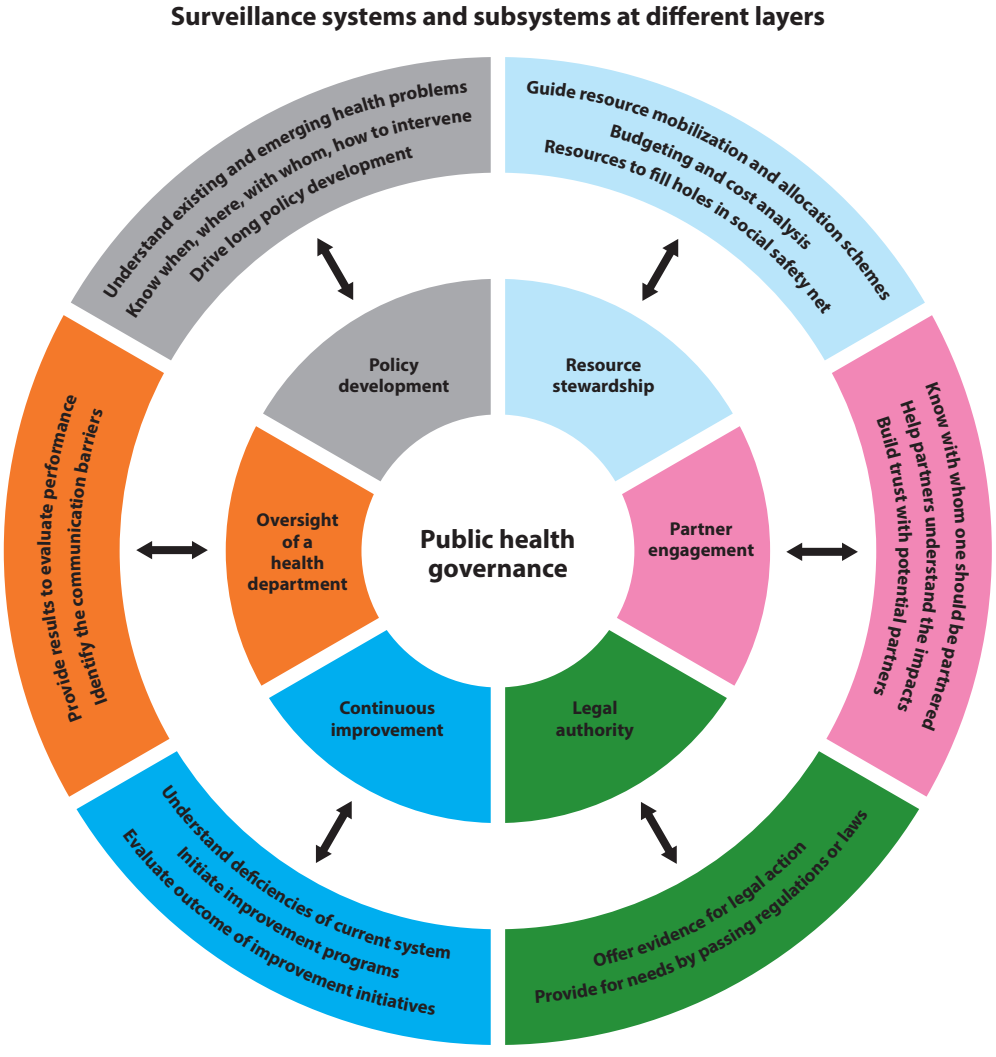


Figure 1

Ideal relationships between public health governance (inner sectors) and surveillance systems (outer sectors).

of which the six core functions are policy development, resource stewardship, continuous improvement, partner engagement, legal authority, and oversight of a health department (**Figure 1**) (12). Achieving efficient and equitable responses and containment strategies also requires improved multilevel collaboration at the global, regional, national, and local levels, particularly with respect to data sharing (24). As shown during the COVID-19 crisis, governance and leadership capacity, along with multilevel coordination, are essential to the strategic deployment of health care resources and the effective mobilization of individuals and communities in pandemic containment (**Figure 2**).

These reforms can shore up system resilience and preparedness, at multiple levels, for pandemics. At its core, control of the COVID-19 pandemic has depended on adequate infrastructure and processes for disease tracking: diagnostics, case identification, and contact tracing. This

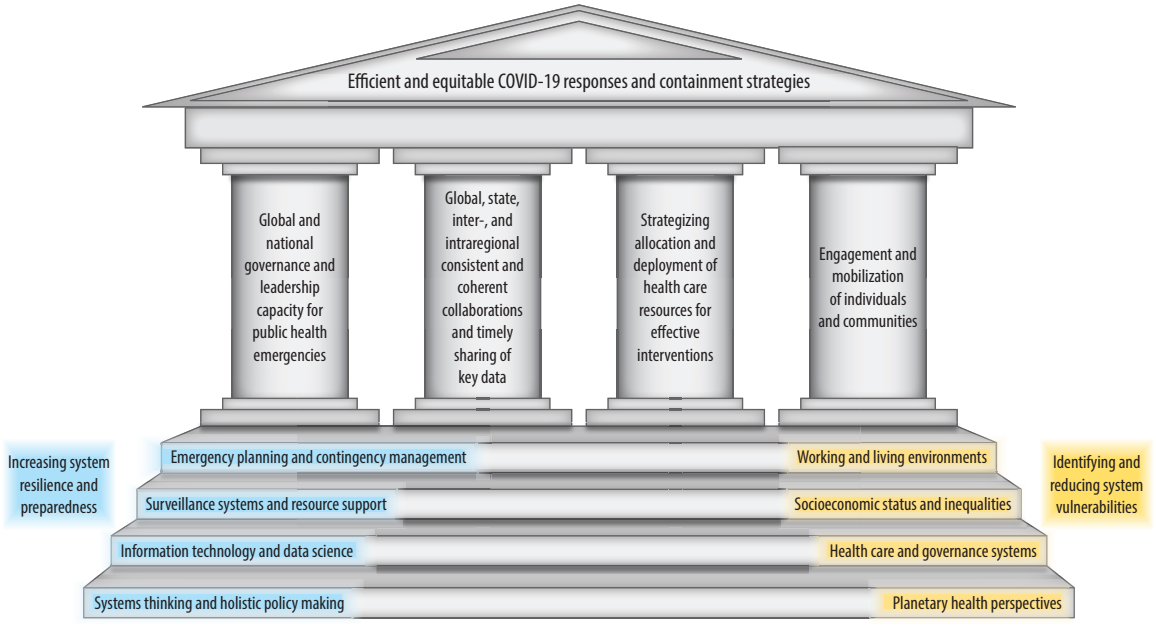


Figure 2

Schematic representation of a systems thinking approach for efficient and equitable COVID-19 responses and containment strategies.

capacity must be supported by a strong public health system that has leveraged other sectors of society through health in all policies (HiAP), which is a collaborative approach that integrates and articulates health considerations into policy making across sectors to improve the health of all communities and people. HiAP recognizes that health is created by a multitude of factors beyond health care and, in many cases, beyond the scope of traditional public health activities (11). In turn, a comprehensive public health response relies on systems thinking, which not only points toward the upstream structural and social determinants of health but also allows for the integration of feedback loops and the consideration of emergent properties.

The tightly coupled elements of interest in public health surveillance (e.g., pathogen, microbial genetic mutation, human socioeconomic activities) demonstrate spatiotemporal dynamics, and their relationships exhibit obvious system properties of adaptive self-organization, being governed by feedback, nonlinearity between effects and actions, historic path dependency, and counterintuitiveness. These features render it barely possible to use traditional methods to conduct public health surveillance. Systems thinking can harness system theories to drive the strategic blueprint and architectural design for a sustainable public health surveillance system, which defines various functions, processes, policies/interventions, governance, and resources required for development and deployment (17, 38, 44, 55, 62). By drawing on system simulation tools, such as system dynamics, agent-based simulation, social network analysis, or any combination thereof (i.e., hybrid models) (7), systems thinking can equip decision makers and policy makers with some pivotal skills, such as performing root cause analyses for problems of interest, scrutinizing critical areas for investigation, conducting system design and refinement, preventing policy backfiring by designing and implementing high-leverage interventions, and enhancing continued learning and improvement from past pandemics (25, 50, 52, 63). Systems thinking also enables a smooth transition of

the data and findings from surveillance systems to the planning, implementation, and evaluation of public health action and provides feedback and guidance on data collection (56).

TECHNOLOGICAL INNOVATIONS IN PUBLIC HEALTH SURVEILLANCE

Information Technology and Data Science

In the early stage of COVID-19 in China, a significant drop in new infections would not have happened in such a short term without the collaborative efforts of many location-aware service providers (i.e., services that provide targeted information to individuals based on their geographic location in real time, such as maps, navigation, and tracing services). Since January 2020, the Chinese mobile phone service providers have been collaborating with the National Health Commission of China, the China CDC, and other relevant governmental sectors and agencies to analyze moving trajectories of COVID-19 cases and their close contacts by cross-referencing location data from multiple sources, including mobile phone service providers, transportation, business transactions (online/offline), resident community screening, neighborhood watch, and social media data (**Figure 3**). China was the first country to develop and put into practice COVID-19-specific apps for tracing and surveillance, with privacy secured (70). Doing so has supplemented the traditional epidemiological survey data by accurately describing the travel history, including daily trajectories, of those infected or those with suspected illness so that close contacts on their trajectories could be identified and notified quickly for self-monitoring and/or isolation from their families to avoid infection being spread. This approach has demonstrated the important emerging roles of information technology and big data in epidemic control and prevention. For example, flight-booking data and intercity human mobility data have been used to estimate the spread of COVID-19 (40, 87); smartphone-based tracing data and social media data have been used to understand spreading mechanisms and predict pandemic trends (46, 47, 66, 82). Combining epidemiological surveys with app-based tracing and derived geo-information has helped to strengthen spatial data infrastructure for infectious surveillance systems and minimize the impact of COVID-19 (67, 71, 97).

Despite significant technological progress, the theories and practices of infectious disease epidemiology and public health surveillance have struggled to keep pace with the changing nature of epidemics and pandemics in the twenty-first century (4). As mentioned above, testing and contact tracing are cornerstones of timely and effective response. While much has been said and written about big data and the development of artificial intelligence, data infrastructure has room for improvement in terms of the timely collection and analysis of cases and related social determinants. Information technology and surveillance systems could harness these new technologies to track not only infections, but also social, economic, and environmental living conditions to address inequalities in risk (96). The rapid development of location-based technologies, including GIS-facilitated data collection, analysis, and visualization, has made important contributions to both global and local action (28, 34, 90). Remote sensing, featuring the simultaneous data acquisition capacity over a large region and a short revisit time over the same location on Earth, can monitor the impact of environmental changes (e.g., global warming) on infectious disease risk (31, 32). Such big data unbounded by administrative boundaries could also drive estimation beyond administrative boundaries. In addition, in the future, lab test results and contact tracing data can be further integrated through mobile devices, which could address test result deduplication issues by, for example, linking together all lab results on the same person over time and across systems, making contact tracing more effective. However, data-driven decision-making is only as comprehensive as the data or evidence it contains. Therefore, the means by which and extent to which such

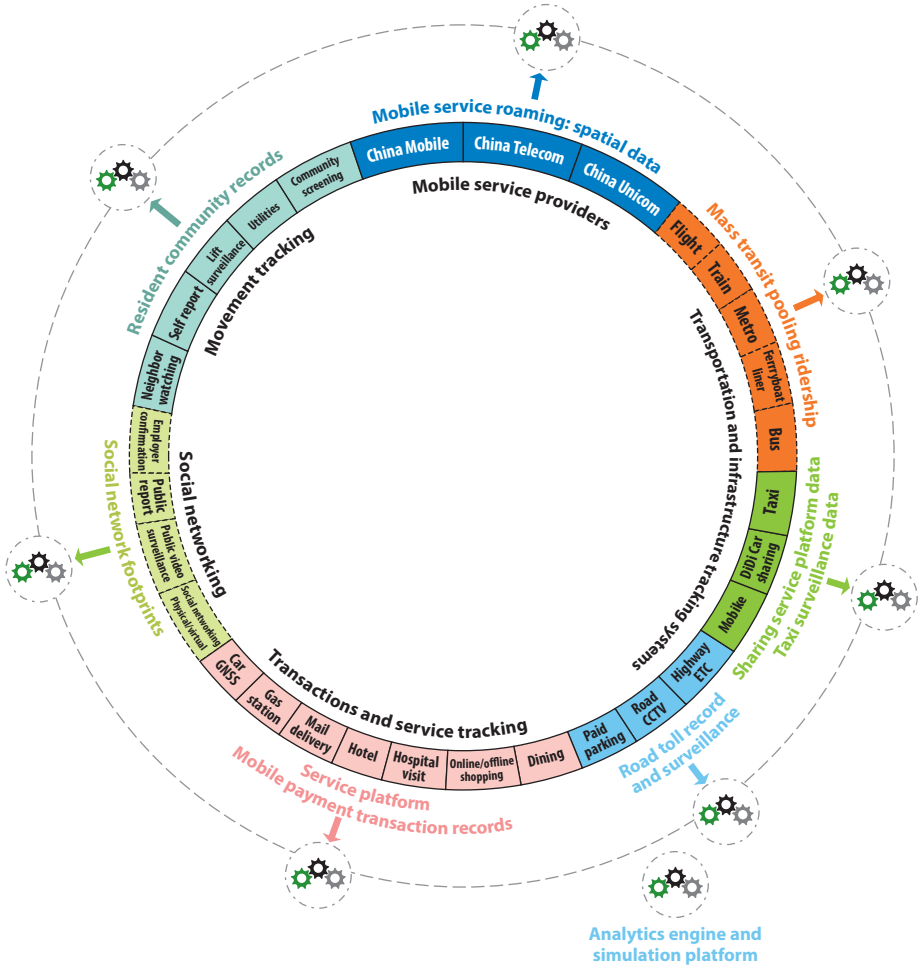


Figure 3

Digital twin used for tracking dynamic risk for the COVID-19 pandemic. Abbreviations: CCTV, closed-circuit television; ETC, electronic toll collection; GNSS, global navigation satellite systems.

technologies are applied are likely to be highly context specific and, in many cases, may even be difficult to adopt because of concerns regarding individual privacy.

The systematic monitoring of interventions, and people's perception, attitudes, daily movements, social interactions, and physical and mental health, can provide valuable insights to guide future actions. Thus, the variability in COVID-19 contexts and response policies presents many opportunities for natural experiments (76). Such underutilized opportunities call for improved population-based longitudinal studies that assess both structural determinants of health as well as individual behavior and health outcomes over time. Location-based technologies and other smartphone-embedded sensors, along with online questionnaire platforms, can help collect such data (which include hard-to-reach populations). Corresponding data infrastructures with near real-time treatment capabilities can provide critical indicators for decision makers. Data privacy issues and other ethical and political challenges relating to the representation of private information in data systems must be transparently and comprehensively addressed to make such efforts

acceptable. Several initiatives are positively contributing to these discussions, including the Global Alliance for Genomics and Health (68) and the Montreal Declaration for Responsible Development of Artificial Intelligence (14). The resulting protocols and policies will be valuable resources for future resilience planning.

Digital Twin: A Robust Tool to Respond to Public Health Emergency Events

Digital twin is the virtual replica of a physical entity or system across its life course and uses real-time big data and other sources to enable visualizing, learning, reasoning, and dynamic recalibration for precision interventions (6, 22). With the acceleration of development in the internet of things—cloud computing, artificial intelligence, big data analytic engines, complex system simulation platforms, and augmented, virtual, and mixed reality—digital twin is witnessing rapid deployment in different industries (16, 19). For example, researchers and engineers in the life science and health care sectors have been exploring the potential applications of integrating genotype and phenotype data with the digital twin to help to develop personalized medicine (60, 69). More examples include in silico genome and digital twins of the human heart and brain, medical equipment, emergency departments, and hospitals (**Figure 4**) (16, 69).

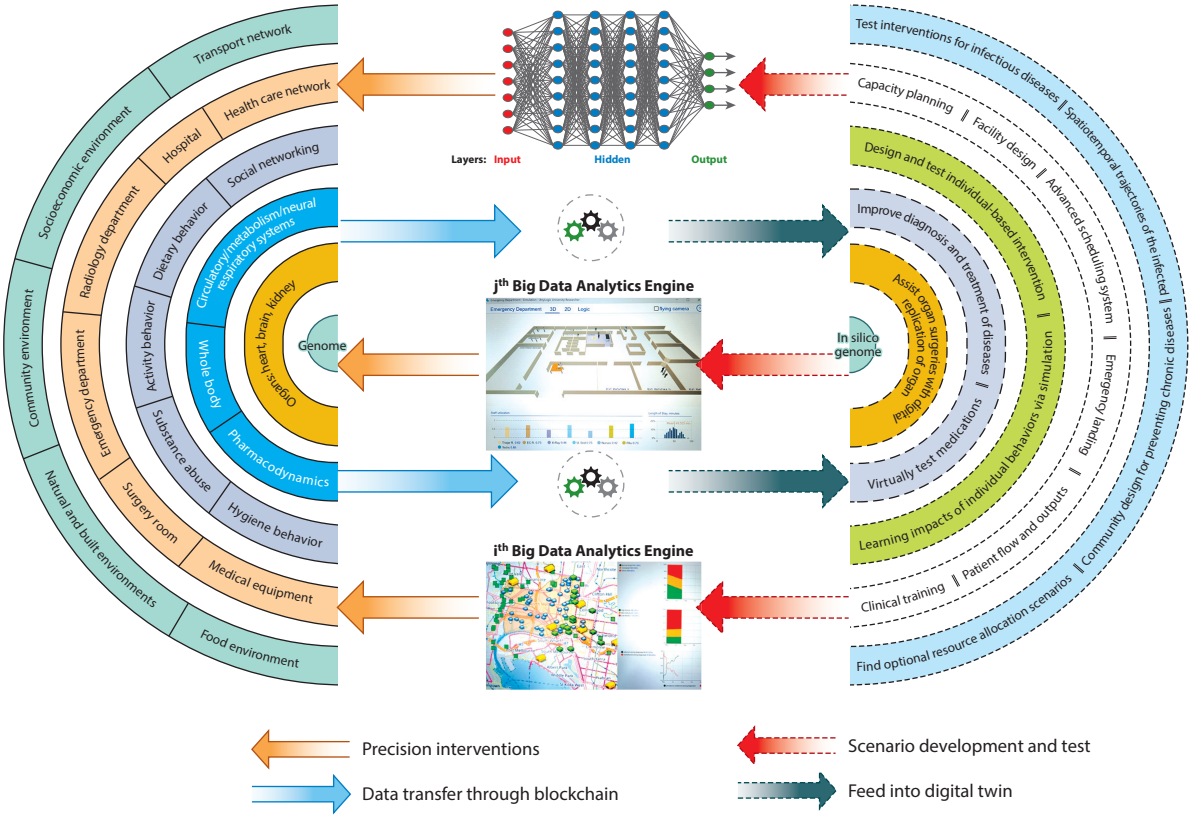


Figure 4

Applications of artificial intelligence, information technology, data science, and digital twin in health care and life science domains. The solid lines that compose the left side semi-circles denote the physical world, and the dashed lines that compose the right side semi-circles denote its corresponding digital twin (i.e., the virtual world).

Digital twin could play important roles at different phases of COVID-19 if its full-fledged potential in public health emergency events is realized. For example, discrete event simulation, agent-based simulation, and virtual reality technologies could be used together to detect flaws in the design of quickly constructed field or cabin hospitals, which may compromise their effectiveness, such as when used for quarantine. Combining internet of things and virtual reality technologies enables health professionals to practice patient care and complicated surgeries ahead of time, which could prevent medical staff from becoming infected while being exposed to patients with extremely high viral loads. By using hybrid models that integrate GIS, agent-based modeling, and social network analysis, researchers could quickly evaluate the spread of infectious diseases and the effectiveness and efficiency of various containment strategies while considering the infrastructure deployment, transport networks and schedules, and individual mobility patterns in virtual reality.

Several concerns exist when applying digital twin. Privacy safety is a major one because this approach involves an intrusive collection of individual location data, which can raise additional ethical issues, such as discrimination (8). Data propriety may also become a concern because agencies might use these data for commercial or other illegal purposes, which can present a risk to regulatory compliance. The vulnerability of the system to hacking and virus attacks is a long-term technical concern, which can result in consequences similar to those from ethical violations. Another concern is how the digital twin is operated: The centralized operation may render the managing agencies as a superpower in terms of controlling citizens' data, while the decentralized operation may significantly reduce the interoperability of different data source

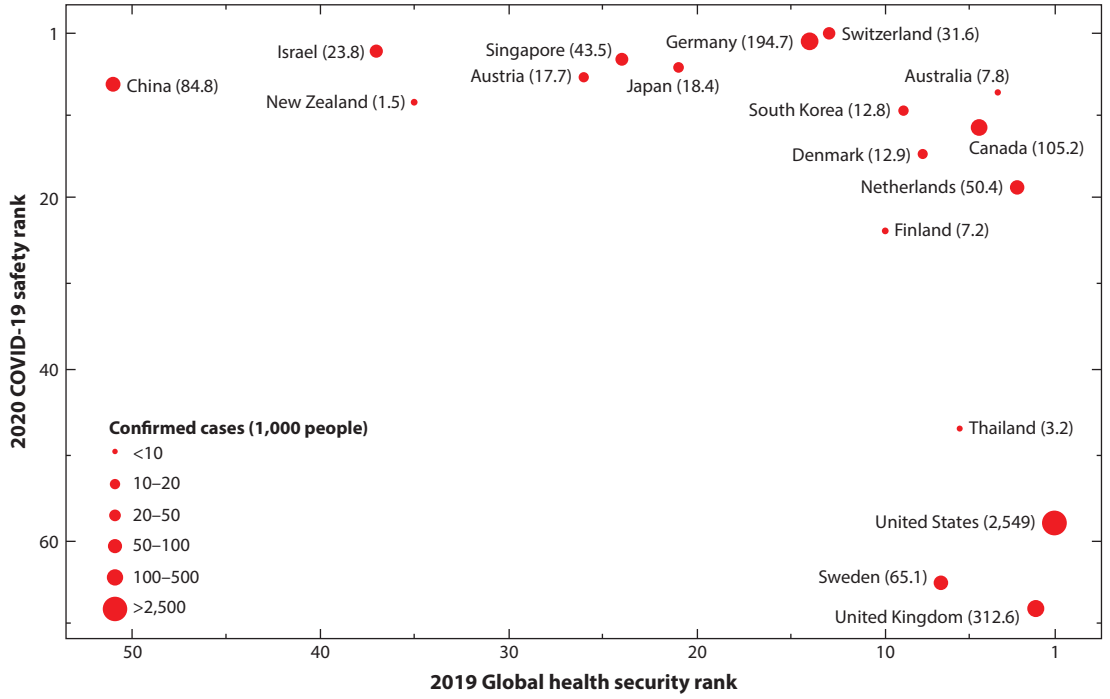


Figure 5

Comparison of the top 10 countries in the 2019 and 2020 health safety rankings (the number in parentheses after each country multiplied by 1,000 is equal to the total number of confirmed COVID-19 infections by June 28, 2020).

providers. However, such privacy-related concerns could be addressed by properly employing new technologies, such as blockchain. Given that digital twin can aid in identifying the vulnerabilities in public health systems and exercise a new way of improving personalized and population health, we suggest that regulations in all countries need to be put in place to ensure that digital twin can be deployed during public health emergencies.

CONCLUDING REMARKS

All the limitations of the current NNDRSs may have hindered the ability to detect the COVID-19 pandemic at the early stage. Even Spain and the United Kingdom, which scored best in terms of real-time public health surveillance and reporting, could not have prevented themselves from becoming the European epicenters of the COVID-19 pandemic. According to the recently published 2020 safety rankings through the months of the pandemic, there have been significant changes in the rankings among countries. These rankings were scored on the basis of quarantine efficiency, monitoring and detection, health readiness, and government efficiency: All countries ranked in the top 10 in 2019 had degraded, with only Australia and South Korea staying at the bottom of the top-10 list in 2020 (**Figure 5**). These findings imply that the ongoing NNDRSs in most if not all countries have not been adequately prepared for pandemic forecasting (10). Technology-driven innovative public health surveillance systems are expected to improve the timeliness, completeness, and accuracy of case reporting during outbreaks and should also enhance feedback and transparency, such that all stakeholders, from public health authorities to the general public, should receive actionable information on infection control and disease risk mitigation earlier than ever before. All epidemic information in one country should consistently stem from one intelligent, robust NNDRS, which should eventually be developed into one global intelligent disease surveillance and reporting system to protect everyone in the world equally.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (42271433), Wuhan University Specific Fund for Major School-Level Internationalization Initiatives (WHU-GJZDZX-PT07), the Ministry of Education Project of Humanities and Social Sciences (21YJAZH053), and the International Institute of Spatial Lifecourse Health (ISLE).

LITERATURE CITED

1. Amit AML, Pepito VCF, Gutierrez B, Rawson T. 2021. Data sharing in Southeast Asia during the first wave of the COVID-19 pandemic. *Front. Public Health* 9:662842
2. Aust. Gov. Dep. Health. 2014. *National framework for communicable disease control*. Rep., Aust. Gov. Dep. Health, Canberra. **https://www.health.gov.au/sites/default/files/documents/2020/10/national-framework-for-communicable-disease-control.pdf**
3. Bagherian H, Farahbakhsh M, Rabiei R, Moghaddasi H, Asadi F. 2017. National Communicable Disease Surveillance System: a review on information and organizational structures in developed countries. *Acta Inform. Med.* 25(4):271–76
4. Bedford J, Farrar J, Ihekweazu C, Kang G, Koopmans M, Nkengasong J. 2019. A new twenty-first century science for effective epidemic response. *Nature* 575(7781):130–36

5. Berger M, Shiau R, Weintraub JM. 2006. Review of syndromic surveillance: implications for waterborne disease detection. *J. Epidemiol. Community Health* 60(6):543–50
6. Bolton RN, McColl-Kennedy JR, Cheung L, Gallan A, Orsingher C, et al. 2018. Customer experience challenges: bringing together digital, physical and social realms. *J. Serv. Manag.* 29(5):776–808
7. Brailsford S, Eldabi T, Kunc M, Mustafee N, Osorio AF. 2019. Hybrid simulation modelling in operational research: a state-of-the-art review. *Eur. J. Oper. Res.* 278(3):721–37
8. Bruynseels K, Santoni de Sio F, van den Hoven J. 2018. Digital twins in health care: ethical implications of an emerging engineering paradigm. *Front Genet*. 9:31
9. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, et al. 2020. Digital technologies in the public-health response to COVID-19. *Nat. Med.* 26:1183–92
10. Cameron EE, Nuzzo JB, Bell JA. 2019. *Global health security index: building collective action and accountability*. Rep., Nucl. Threat Initiat./Johns Hopkins Bloomberg Sch. Public Health, Washington, DC/Baltimore, MD. **https://www.ghsindex.org/wp-content/uploads/2020/04/2019-Global-Health-Security-Index.pdf**
11. CDC (Cent. Dis. Control Prev.). 2016. Health in all policies. *Office of the Associate Director for Policy and Strategy*. **https://www.cdc.gov/policy/hiap/index.html**
12. CDC (Cent. Dis. Control Prev.). 2020. What is NBS? *National Electronic Disease Surveillance System Base System*. **https://www.cdc.gov/nbs/overview/index.html**
13. Chandrasekar K, Mahesan S, Bath PA. 2013. Notifiable disease surveillance in Sri Lanka and the United Kingdom: a comparative study. *Sri Lanka J. Bio-Med. Inform.* 4(1):14–22
14. Coombes MG. 2000. Defining locality boundaries with synthetic data. *Environ. Plan. A Econ. Space* 32(8):1499–518
15. CRS (Congr. Res. Serv.). 2020. *Tracking COVID-19: U.S. public health surveillance and data*. R46588, CRS, Washington, DC. **https://sgp.fas.org/crs/misc/R46588.pdf**
16. DHL Trend Res. 2019. *Digital twins in logistics: a DHL perspective on the impact of digital twins on the logistics industry*. Rep., DHL Trend Res. **https://www.dhl.com/content/dam/dhl/global/core/documents/pdf/glo-core-digital-twins-in-logistics.pdf**
17. Durski KN, Naidoo D, Singaravelu S, Shah AA, Djingarey MH, et al. 2020. Systems thinking for health emergencies: use of process mapping during outbreak response. *BMJ Global Health* 5:e003901
18. Elliot AJ, Harcourt SE, Hughes HE, Loveridge P, Morbey RA, et al. 2020. The COVID-19 pandemic: a new challenge for syndromic surveillance. *Epidemiol. Infect.* 148:e122
19. GE (Gen. Electr.). 2016. *GE digital twin: analytic engine for the digital power plant*. Rep., GE Power Digit. Solut., Boston. **https://www.ge.com/digital/sites/default/files/download_assets/Digital-Twin-for-the-digital-power-plant-.pdf**
20. Gebreyes WA, Dupouy-Camet J, Newport MJ, Oliveira CJB, Schlesinger LS, et al. 2014. The Global One Health paradigm: challenges and opportunities for tackling infectious diseases at the human, animal, and environment interface in low-resource settings. *PLOS Negl. Trop. Dis.* 8(11):e3257
21. George J, Häsler B, Mremi I, Sindato C, Mboera L, et al. 2020. A systematic review on integration mechanisms in human and animal health surveillance systems with a view to addressing global health security threats. *One Health Outlook* 2:11
22. Grieves M. 2019. Virtually intelligent product systems: digital and physical twins. In *Complex Systems Engineering: Theory and Practice*, ed. S Flumerfelt, KG Schwartz, D Mavris, S Briceno, pp. 175–200. Reston, VA: Am. Inst. Aeronaut. Astronaut.
23. Groseclose SL, Buckeridge DL. 2017. Public health surveillance systems: recent advances in their use and evaluation. *Annu. Rev. Public Health* 38:57–79
24. Haffajee RL, Mello MM. 2020. Thinking globally, acting locally—the U.S. response to Covid-19. *N. Engl. J. Med.* 382(22):e75
25. Hassan I, Obaid F, Ahmed R, Abdelrahman L, Adam S, et al. 2020. A systems thinking approach for responding to the COVID-19 pandemic. *East Mediterr. Health J.* 26(8):872–76
26. Holmgren AJ, Apathy NC, Adler-Milstein J. 2020. Barriers to hospital electronic public health reporting and implications for the COVID-19 pandemic. *J. Am. Med. Inf. Assoc.* 27(8):1306–9
27. Jayasinghe S. 2015. Social determinants of health inequalities: towards a theoretical perspective using systems science. *Int. J. Equity Health* 14(1):71

28. Jia P. 2020. Understanding the epidemic course in order to improve epidemic forecasting. *Geohealth* 4(10):e2020GH000303
29. Jia P. 2021. A changed research landscape of youth's obesogenic behaviours and environments in the post-COVID-19 era. *Obes. Rev.* 22(Suppl. 1):e13162
30. Jia P, Dai S, Wu T, Yang S. 2021. New approaches to anticipate the risk of reverse zoonosis. *Trends Ecol. Evol.* 36(7):580–90
31. Jia P, Stein A, James P, Brownson RC, Wu T, et al. 2019. Earth observation: investigating noncommunicable diseases from space. *Annu. Rev. Public Health* 40:85–104
32. Jia P, Wang T, van Vliet AJH, Skidmore AK, van Aalst M. 2020. Worsening of tree-related public health issues under climate change. *Nat. Plants* 6(2):48
33. Jia P, Yang S. 2020. China needs a national intelligent syndromic surveillance system (NISSS). *Nat. Med.* 26:990
34. Jia P, Yang S. 2020. Early warning of epidemics: towards a national intelligent syndromic surveillance system (NISSS) in China. *BMJ Glob. Health* 5(10):e002925
35. Jia P, Zhang L, Yu W, Yu B, Liu M, et al. 2021. Impact of COVID-19 lockdown on activity patterns and weight status among youths in China: the COVID-19 Impact on Lifestyle Change Survey (COINLICS). *Int. J. Obes.* 45:695–99
36. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, et al. 2008. Global trends in emerging infectious diseases. *Nature* 451(7181):990–93
37. Jones T, Hedberg C. 2018. Coordination of multiple states and federal agencies. *CDC Field Epidemiology Manual*. **https://www.cdc.gov/eis/field-epi-manual/chapters/Coordinating-Agencies.html**
38. Katapally TR. 2019. The SMART framework: integration of citizen science, community-based participatory research, and systems science for population health science in the digital age. *JMIR mHealth uHealth* 7(8):e14056
39. Kostkova P. 2018. Disease surveillance data sharing for public health: the next ethical frontiers. *Life Sci. Soc. Policy* 14(1):16
40. Kraemer MUG, Yang C-H, Gutierrez B, Wu C-H, Klein B, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368(6490):493–97
41. Lancet. 2020. COVID-19: remaking the social contract. *Lancet* 395(10234):1401
42. Lawpoolsri S, Kaewkungwal J, Khamsiriwatchara A, Sovann L, Sreng B, et al. 2018. Data quality and timeliness of outbreak reporting system among countries in greater Mekong subregion: challenges for international data sharing. *PLOS Negl. Trop. Dis.* 12:e0006425
43. Lee D, Robertson C, Marques D. 2022. Quantifying the small-area spatio-temporal dynamics of the Covid-19 pandemic in Scotland during a period with limited testing capacity. *Spat Stat*. 49:100508
44. Leischow SJ, Best A, Trochim WM, Clark PI, Gallagher RS, et al. 2008. Systems thinking to improve the public's health. *Am. J. Prev. Med.* 35(S2):S196–203
45. Leitmeyer KC, Espinosa L, Broberg EK, Struelens MJ, ECDC Natl. Focal Points Lab. 2020. Automated digital reporting of clinical laboratory information to national public health surveillance systems, results of a EU/EEA survey, 2018. *Euro. Surveill.* 25(39):1900591
46. Leung K, Wu JT, Leung GM. 2021. Real-time tracking and prediction of COVID-19 infection using digital proxies of population mobility and mixing. *Nat. Commun.* 12:1501
47. Liang F. 2020. COVID-19 and health code: how digital platforms tackle the pandemic in China. *Soc. Media Soc.* 6(3). **https://doi.org/10.1177/2056305120947657**
48. Limkittikul K, Brett J, L'Azou M. 2014. Epidemiological trends of dengue disease in Thailand (2000–2011): a systematic literature review. *PLOS Negl. Trop. Dis.* 8(11):e3241
49. Liverani M, Teng S, Le MS, Coker R. 2018. Sharing public health data and information across borders: lessons from Southeast Asia. *Glob. Health* 14(1):94
50. Luke DA, Stamatakis KA. 2012. Systems science methods in public health: dynamics, networks, and agents. *Annu. Rev. Public Health* 33:357–76
51. Luo M, Wang Q, Yang S, Jia P. 2022. Changes in patterns of take-away food ordering among youths before and after COVID-19 lockdown in China: the COVID-19 Impact on Lifestyle Change Survey (COINLICS). *Eur. J. Nutr.* 61:1121–31

52. Mabry PL, Marcus SE, Clark PI, Leischow SJ, Méndez D. 2010. Systems science: a revolution in public health policy research. *Am. J. Public Health* 100(7):1161–63
53. Maxmen A. 2020. Why the United States is having a coronavirus data crisis. *Nature* 585:13–14
54. McCloskey B, Heymann DL. 2020. SARS to novel coronavirus—old lessons and new lessons. *Epidemiol. Infect.* 148:e22
55. McNab D, McKay J, Shorrock S, Luty S, Bowie P. 2020. Development and application of 'systems thinking' principles for quality improvement. *BMJ Open. Qual.* 9:e000714
56. McNabb SJN, Chungong S, Ryan M, Wuhib T, Nsubuga P, et al. 2002. Conceptual framework of public health surveillance and action and its application in health sector reform. *BMC Public Health* 2:2
57. Meyers L, Ginocchio CC, Faucett AN, Nolte FS, Gesteland PH, et al. 2018. Automated real-time collection of pathogen-specific diagnostic data: syndromic infectious disease epidemiology. *JMIR Public Health Surveill.* 4(3):e59
58. Osaka K, Takahashi H, Ohyama T. 2002. Testing a symptom-based surveillance system at high-profile gatherings as a preparatory measure for bioterrorism. *Epidemiol. Infect.* 129(3):429–34
59. Park S, Cho E. 2014. National Infectious Diseases Surveillance data of South Korea. *Epidemiol. Health* 36:e2014030
60. Persistent Syst. 2022. *Digital twins in healthcare: enabling mass personalization of care delivery*. White Pap., Persistent Syst., New York. **https://www.persistent.com/insights/whitepapers/digital-twins-in-healthcare/**
61. Peters DH. 2014. The application of systems thinking in health: Why use systems thinking? *Health Res. Policy Sys.* 12:51
62. Plack MM, Goldman EF, Scott AR, Brundage SB. 2019. *Systems Thinking in the Healthcare Professions: A Guide for Educators and Clinicians*. Washington, DC: George Wash. Univ. **https://hsrc.himmelfarb.gwu.edu/cgi/viewcontent.cgi?article=1000&context=educational_resources_teaching**
63. Rahmandad H, Sterman JD. 2008. Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models. *Manag. Sci.* 54(5):998–1014
64. Reskin B. 2012. The race discrimination system. *Annu. Rev. Sociol.* 38:17–35
65. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. 2015. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin. Epidemiol.* 7:449–90
66. Shahroz M, Ahmad F, Younis MS, Ahmad N, Kamel Boulos MNK, et al. 2021. COVID-19 digital contact tracing applications and techniques: a review post initial deployments. *Transp. Eng.* 5:100072
67. Shi Y, Jiang H-L, Yang M-X, Dong L-J, Chen Y, et al. 2021. The precision of epidemiological investigation of COVID-19 transmission in Shanghai, China. *Infect. Dis. Poverty* 10:58
68. Shortt NK, Moore A, Coombes M, Wymer C. 2005. Defining regions for locality health care planning: a multidimensional approach. *Soc. Sci. Med.* 60(12):2715–27
69. Siemens. *The value of digital twin technology: transforming care delivery in a single department or an entire healthcare enterprise*. White Pap., Siemens, Erlangen, Ger. **https://www.siemens-healthineers.com/en-us/services/value-partnerships/asset-center/white-papers-articles/value-of-digital-twin-technology**
70. Singh HJL, Couch D, Yap K. 2020. Mobile health apps that help with COVID-19 management: scoping review. *JMIR Nurs.* 3(1):e20596
71. Smith CD, Mennis J. 2020. Incorporating geographic information science and technology in response to the COVID-19 pandemic. *Prev. Chronic Dis.* 17:E58
72. Smith GE, Elliot AJ, Lake I, Edeghere O, Morbey R, et al. 2019. Syndromic surveillance: two decades experience of sustainable systems—its people not just data! *Epidemiol. Infect.* 147:e101
73. Smith T, Rotondo J, Desai S, Deehan H. 2014. Pertussis surveillance in Canada: trends to 2012. *Can. Commun. Dis. Rep.* 40(3):21–30
74. Statens Serum Inst. 2016. Surveillance in Denmark. *Statens Serum Institut*. **https://en.ssi.dk/surveillance-and-preparedness/surveillance-in-denmark**
75. Thacker SB, Berkelman RL. 1988. Public health surveillance in the United States. *Epidemiol. Rev.* 10:164–90

76. Thomson B. 2020. The COVID-19 pandemic: a global natural experiment. *Circulation* 142:14–16
77. Thygesen LC, Daasnes C, Thaulow I, Brønnum-Hansen H. 2011. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand. J. Public Health* 39(Suppl. 7):12–16
78. Tosti ME, Longhi S, de Waure C, Mele A, Frano E, et al. 2015. Assessment of timeliness, representativeness and quality of data reported to Italy's national integrated surveillance system for acute viral hepatitis (SEIEVA). *Public Health* 129(5):561–68
79. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, et al. 2014. A systematic review of barriers to data sharing in public health. *BMC Public Health* 14:1144
80. Vlieg WL, Fanoy EB, van Asten L, Liu X, Yang J, et al. 2017. Comparing national infectious disease surveillance systems: China and the Netherlands. *BMC Public Health* 17(1):415
81. Wakefield J, Okonek T, Pedersen J. 2020. Small area estimation for disease prevalence mapping. *Int. Stat. Rev.* 88(2):398–418
82. Wang S, Ding S, Xiong L. 2020. A new system for surveillance and digital contact tracing for COVID-19: spatiotemporal reporting over network and GPS. *JMIR mHealth uHealth* 8(6):e19457
83. WHO (World Health Organ.). 2020. *Public health surveillance for COVID-19*. Interim Guid. 2020.8, WHO, Geneva. **https://apps.who.int/iris/bitstream/handle/10665/337897/WHO-2019-nCoV-SurveillanceGuidance-2020.8-eng.pdf**
84. Wilder-Smith A, Freedman DO. 2020. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *J. Travel. Med.* 27(2):taaa020
85. Wójcik OP, Brownstein JS, Chunara R, Johansson MA. 2014. Public health for the people: participatory infectious disease surveillance in the digital age. *Emerg. Themes Epidemiol.* 11(1):7
86. Wood A, Denholm R, Hollings S, Cooper J, Ip S, et al. 2021. Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource. *BMJ* 373:n826
87. Wu JT, Leung K, Leung GM. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 395(10225):689–97
88. Wu Z, McGoogan JM. 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 323(13):1239–42
89. Yang S, Pan X, Zeng P, Jia P. 2021. Spatial technologies to strengthen traditional testing for SARS-CoV-2. *Trends Microbiol.* 29(12):1055–57
90. Yang S, Yu C, Jia P. 2021. Spatiobehavioral characteristics—defining the epidemiology of new contagious diseases at the earliest moment possible. *Trends Parasitol.* 37(3):179–81
91. Yoo H-S, Park O, Park H-K, Lee E-G, Jeong E-K, et al. 2009. Timeliness of national notifiable diseases surveillance system in Korea: a cross-sectional study. *BMC Public Health* 9(1):93
92. Yu B, Luo M, Liu M, Zhou J, Yang S, Jia P. 2021. Social capital changes after COVID-19 lockdown among youths in China: COVID-19 Impact on Lifestyle Change Survey (COINLICS). *Front. Public Health* 9:697068
93. Yu B, Zhang D, Yu W, Luo M, Yang S, Jia P. 2021. Impacts of lockdown on dietary patterns among youths in China: the COVID-19 Impact on Lifestyle Change Survey. *Public Health Nutr.* 24(11):3221–32
94. Zhang L, Zhang L, Lai L, Du Z, Huang Y, et al. 2022. Risk assessment of imported COVID-19 in China: a modelling study in Sichuan Province. *Transbound. Emerg. Dis.* **https://doi.org/10.1111/tbed.14700**. In press
95. Zhang X, Holt JB, Yun S, Lu H, Greenlund KJ, Croft JB. 2015. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the Behavioral Risk Factor Surveillance System. *Am. J. Epidemiol.* 182(2):127–37
96. Zhang Y, Peng N, Yang S, Jia P. 2022. Associations between nighttime light and COVID-19 incidence and mortality in the United States. *Int. J. Appl. Earth Obs. Geoinf.* 112:102855
97. Zhou Y, Jiang H, Wang Q, Yang M, Chen Y, Jiang Q. 2021. Use of contact tracing, isolation, and mass testing to control transmission of covid-19 in China. *BMJ* 375:n2330

Annual Review of Public Health
Volume 44, 2023

Contents

Epidemiology and Biostatistics

Social Environment and Behavior

Environmental and Occupational Health

Public Health Practice and Policy

Health Services

Indexes

Errata

An online log of corrections to *Annual Review of Public Health* articles may be found at http://www.annualreviews.org/errata/publhealth

Related Articles

From the ***Annual Review of Animal Biosciences***, Volume 11 (2023)

Animal Models, Zoonotic Reservoirs, and Cross-Species Transmission of Emerging Human-Infecting Coronaviruses
Yakhouba Kane, Gary Wong, and George F. Gao

Domestic Animals as Potential Reservoirs of Zoonotic Viral Diseases
Oyewale Tomori and Daniel O. Oluwayelu

From the ***Annual Review of Biomedical Data Science***, Volume 5 (2022)

Genome Privacy and Trust
Gamze Gürsoy

Developing and Implementing Predictive Models in a Learning Healthcare System: Traditional and Artificial Intelligence Approaches in the Veterans Health Administration
David Atkins, Christos A. Makridis, Gil Alterovitz, Rachel Ramoni, and Carolyn Clancy

From the ***Annual Review of Cancer Biology***, Volume 6 (2022)

Tracing and Targeting the Origins of Childhood Cancer
Tim H.H. Coorens and Sam Behjati

From the ***Annual Review of Clinical Psychology***, Volume 18 (2022)

Cognitive Aging and the Promise of Physical Activity
Kirk I. Erickson, Shannon D. Donofry, Kelsey R. Sewell, Belinda M. Brown, and Chelsea M. Stillman

Police Violence and Public Health
Jordan E. DeVylder, Deidre M. Anglin, Lisa Bowleg, Lisa Fedina, and Bruce G. Link

The Psychology of Pandemics
Steven Taylor

From the ***Annual Review of Criminology***, Volume 6 (2023)

The Opioid Crisis: The War on Drugs Is Over. Long Live the War on Drugs
Marie Gottschalk

From the ***Annual Review of Developmental Psychology***, Volume 4 (2022)

Practice and Policy Regarding Child Neglect: Lessons from Studies of Institutional Deprivation
Charles H. Zeanah and Lucy S. King

From the ***Annual Review of Economics***, Volume 14 (2022)

The Impact of Health Information and Communication Technology on Clinical Quality, Productivity, and Workers
Ari Bronsoler, Joseph Doyle, and John Van Reenen

The Economics of the COVID-19 Pandemic in Poor Countries
Edward Miguel and Ahmed Mushfiq Mobarak

The Affordable Care Act After a Decade: Industrial Organization of the Insurance Exchanges
Benjamin Handel and Jonathan Kolstad

From the ***Annual Review of Environment and Resources***, Volume 47 (2022)

COVID-19 and the Environment: Short-Run and Potential Long-Run Impacts
Noah S. Diffenbaugh

Sustainability in Health Care
Howard Hu, Gary Cohen, Bhavna Sharma, Hao Yin, and Rob McConnell

Agrochemicals, Environment, and Human Health
P. Indira Devi, M. Manjula, and R.V. Bhavani

The Concept of Adaptation
Ben Orlove

From the ***Annual Review of Law and Social Science***, Volume 18 (2022)

Good Law to Fight Bad Bugs: Legal Responses to Epidemics
Carol A. Heimer and Clay Davis

Environmental Legal Mobilization
Lisa Vanhala

From the ***Annual Review of Medicine***, Volume 74 (2023)

COVID-19: Challenges of Viral Variants
Jana L. Jacobs, Ghady Haidar, and John W. Mellors

Post-COVID-19 Condition
Ani Nalbandian, Amar D. Desai, and Elaine Y. Wan

Maternal Mortality in the United States: Trends and Opportunities for Prevention
Siwen Wang, Kathryn M. Rexrode, Andrea A. Florio, Janet W. Rich-Edwards, and Jorge E. Chavarro

Diverse Approaches to Gene Therapy of Sickle Cell Disease
Shanna L. White, Kevyn Hart, and Donald B. Kohn

From the ***Annual Review of Nutrition***, Volume 42 (2022)

The Importance of Food Processing and Eating Behavior in Promoting Healthy and Sustainable Diets
Ciarán G. Forde and Eric A. Decker

Folic Acid and the Prevention of Birth Defects: 30 Years of Opportunity and Controversies
Krista S. Crider, Yan Ping Qi, Lorraine F. Yeung, Cara T. Mai, Lauren Head Zauche, Arick Wang, Kelicia Daniels, and Jennifer L. Williams

Advancing Health Equity Efforts to Reduce Obesity: Changing the Course
Shiriki K. Kumanyika

From the ***Annual Review of Political Science***, Volume 25 (2022)

Three Faces of Climate Justice
Nives Dolšak and Aseem Prakash

Media and Policy Making in the Digital Age
Emiliano Grossman

From the ***Annual Review of Psychology***, Volume 74 (2023)

Psychology of Climate Change
Linda Steg

From the ***Annual Review of Statistics and Its Application***, Volume 9 (2022)

Is There a Cap on Longevity? A Statistical Review
Léo R. Belzile, Anthony C. Davison, Jutta Gampe, Holger Rootzén, and Dmitrii Zholud

Framing Causal Questions in Life Course Epidemiology
Bianca L. De Stavola, Moritz Herle, and Andrew Pickles

From the ***Annual Review of Virology***, Volume 9 (2022)

Lessons from Acquired Natural Immunity and Clinical Trials to Inform Next-Generation Human Cytomegalovirus Vaccine Development
Xintao Hu, Hsuan-Yuan Wang, Claire E. Otero, Jennifer A. Jenks, and Sallie R. Permar

Data challenges for international health emergencies: lessons learned from ten international COVID-19 driver projects

Sally Boylan, Catherine Arsenault, Marcos Barreto, Fernando A Bozza, Adalton Fonseca, Eoghan Forde, Lauren Hookham, Georgina S Humphreys, Maria Yury Ichihara, Kirsty Le Doare, Xiao Fan Liu, Edel McNamara, Jean Claude Mugunga, Juliane F Oliveira, Joseph Ouma, Neil Postlethwaite, Matthew Retford, Luis Felipe Reyes, Andrew D Morris, Anne Wozencraft

The COVID-19 pandemic highlighted the importance of international data sharing and access to improve health outcomes for all. The International COVID-19 Data Alliance (ICODA) programme enabled 12 exemplar or driver projects to use existing health-related data to address major research questions relating to the pandemic, and developed data science approaches that helped each research team to overcome challenges, accelerate the data research cycle, and produce rapid insights and outputs. These approaches also sought to address inequity in data access and use, test approaches to ethical health data use, and make summary datasets and outputs accessible to a wider group of researchers. This Health Policy paper focuses on the challenges and lessons learned from ten of the ICODA driver projects, involving researchers from 19 countries and a range of health-related datasets. The ICODA programme reviewed the time taken for each project to complete stages of the health data research cycle and identified common challenges in areas such as data sharing agreements and data curation. Solutions included provision of standard data sharing templates, additional data curation expertise at an early stage, and a trusted research environment that facilitated data sharing across national boundaries and reduced risk. These approaches enabled the driver projects to rapidly produce research outputs, including publications, shared code, dashboards, and innovative resources, which can all be accessed and used by other research teams to address global health challenges.

Lancet Digit Health 2024; 6: e354–66

Health Data Research UK, London, UK (S Boylan MSc, E McNamara LLM, N Postlethwaite BSc, M Retford BCompSci, Prof A D Morris MD, A Wozencraft PhD)**; Department of Global Health, Milken Institute School of Public Health, George Washington University, Washington, DC, USA** (C Arsenault PhD)**; Center for Data and Knowledge Integration for Health, Gonçalo Moniz Institute, Oswaldo Cruz Foundation, Salvador, Brazil** (M Barreto PhD, A Fonseca PhD, M Y Ichihara PhD, J F Oliveira PhD)**; Evandro Chagas National Institute of Infectious Disease, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil** (F A Bozza PhD MD)**; Aridhia Informatics, Glasgow, UK** (E Forde PhD)**; St George's, University of London, London, UK** (L Hookham MBBS, Prof K Le Doare PhD)**; Green Templeton College** (G S Humphreys PhD) **and Nuffield School of Medicine** (L F Reyes PhD MD)**, University of Oxford, Oxford, UK; Makerere University John's Hopkins University Research Collaboration, Kampala, Uganda** (Prof K Le Doare, J Ouma PhD)**; Department of Media and Communication, City University of Hong Kong, Hong Kong Special Administrative Region, China** (X F Liu PhD)**; Partners in Health, Boston, MA, USA** (J C Mugunga MD MS)**; Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA** (J C Mugunga)**; Division of Global Health Equity, Brigham and Women's Hospital, Boston, MA, USA** (J C Mugunga)**; Department of Mathematics, Centre of Mathematics of the University of Porto, Porto,**

Introduction

Data are at the core of patient care, population health management, health-service planning, and research. Amid the loss of health and life to COVID-19, researchers and policy makers engaged with data and information more intensively than ever, but major challenges to international data access and sharing were exposed. These challenges led the G7 nations to prioritise the establishment of health data as a global public good.[1] As such, the COVID-19 pandemic highlighted the importance of effective, equitable, and ethical data sharing, and the power of timely data sharing to provide crucial new insights and improve health-care outcomes.[2] Initiatives and systems to enable sharing of health-relevant data were established,[3] and many researchers, particularly in lockdowns, focused their efforts on maximising the benefits from secondary data.[4] In contrast to primary data collection, secondary data refers to information collected for purposes other than the intended research.[5] These data can come from a wide range of pre-existing sources including disease registries, health management and planning systems, clinical care records, and epidemiological surveillance tools. Due to the broad spectrum of secondary data sources, accessing, preparing, and analysing these data for research purposes presents numerous and distinct challenges.[6]

The International COVID-19 Data Alliance (ICODA) programme was convened by Health Data Research (HDR) UK in July, 2020. Its aim was to make existing, health-relevant research data accessible to researchers everywhere, enabling them to address key questions relating to COVID-19 and provide new insights that led to improved health outcomes for all, with a particular focus on low-income and middle-income countries. To achieve this vision, it assembled an open, international alliance of partners that brought together stakeholders, including community and patient representatives, to shape the programme and show trustworthiness. This approach built on that of other initiatives that have brought together international communities of health-care and research organisations to develop agreed standards and frameworks for the ethical use of health data to address global health challenges, such as the Global Alliance for Genomics and Health[7] and the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC).[8]

The ICODA initiative made use of the Five Safes framework,[9] first developed by the UK's Office for National Statistics and adopted by HDR UK, as an approach to enable the right to privacy while unlocking the power of data for research (panel 1). ICODA focused on improving data discoverability through the ICODA Gateway, a dataset catalogue and access tool, and on providing researchers with a trusted research environment (TRE) called the ICODA Workbench, a highly secure and controlled computing environment that allowed approved researchers from authorised organisations a safe way to access, store, and analyse sensitive data remotely.[12] Working in partnership with data curation and databank providers, the Workbench enabled secure data access and analysis for data partners and researchers from over 70 countries, as well as a collaborative space for their projects.

Central to ICODA's approach was a cohort of 12 exemplar or driver projects, for which the core team

Key messages

- Appropriate data infrastructure, governance, and support should be provided early, to enable insights to be generated rapidly.
- Standard data sharing agreements and templates can speed up what can typically be a lengthy process.
- Use of pre-provisioned trusted research environments can go a long way to opening up data sharing across national and regional boundaries; expediting this process can be crucial in research areas such as rare diseases, where national datasets might be too small to give rise to significant results. It also provides a good mechanism for reducing the risk involved in data sharing, as the data remains within a secure environment at all times.
- Use of data curation expertise early on in initiatives can accelerate progress as this step is typically time-consuming and often underestimated. As part of this curation, considering making data findable, accessible, interoperable, and reusable at the same time and considering field labelling and units can reduce the work involved in sharing metadata.
- Community and wider stakeholder engagement requires an investment of time and resources, but is crucial to building trust and ensuring that relevant research questions are addressed and insights taken up into policy and practice.
- The Five Safes framework resonates with a global audience and provides an understandable and readily accessible framework with broad applicability.
- Willingness to conduct open science is crucial not only to making datasets discoverable to other researchers, but also to share code, methods, and lessons learned. It can also improve trust and transparency and, in future, promote easier sharing of data.
- Bringing together cohorts of researchers, even if working on disparate research questions, results in synergy and rapid identification of problems that need to be resolved for multiple parties.

Panel 1: International COVID-19 Data Alliance (ICODA) implementation of the Five Safes framework

(1) Safe people: ICODA implemented a proportionate researcher accreditation process to ensure data are only accessed by those who are trained and trusted to use it appropriately[10]
(2) Safe settings: a secure trusted research environment, the ICODA Workbench, was provided for use by driver projects
(3) Safe projects: the Grand Challenges ICODA open funding call review identified projects with rigorous project management and delivery plans in place
(4) Safe data: teams were supported in ensuring project data were de-identified
(5) Safe outputs: an output review process was developed to ensure outputs were non-disclosive and, where appropriate, validated the scientific integrity of results[11]

Portugal (J F Oliveira); **Universidad de La Sabana, Chia, Colombia** (L F Reyes)

Correspondence to: Sally Boylan, Health Data Research UK, London NW1 2BE, UK
sally.boylan@hdruk.ac.uk

For more on **ICODA** see https://icoda-research.org/about/about-us/

For more on the **ICODA Workbench** see https://icoda-research.org/research/our-research/

For more on the **ICODA driver projects** see https://icoda-research.org/research/driver-projects/

worked in close partnership with data contributors and the research community. Each driver project addressed important COVID-19-related research questions and ran at the same time as the ICODA programme was being developed, serving to test and shape ICODA's processes and tools as overall approaches were being developed for the long term. The projects allowed ICODA to bring together data, make them accessible, and generate valuable insights. The initial driver projects, Efficacy and Safety of COVID-19 Treatments and International Perinatal Outcomes in the Pandemic, were the first to co-create and test these approaches and provided valuable learning that benefited subsequent driver projects.[13,14]

Following an international call for proposals, a further ten driver projects were identified and established in July, 2021, through 12-month Grand Challenges ICODA research awards, and these are the main focus of this Health Policy (table). These projects addressed major research questions relating to the COVID-19 pandemic, such as the effectiveness of community-based vaccination programmes and the effect of the pandemic on health-service delivery. All used innovative data science approaches applied to existing health data and aimed to help address global inequity in access, the quality and use of data, test new approaches to support ethical and trustworthy health data use, and deliver rapid outputs and insights. Some of the projects involved large teams and datasets from multiple countries, with many researchers based in low-income and middle income countries (figure 1).

Despite the many challenges that remain in health data reuse, the ICODA programme supported 135 researchers from 19 countries to access and analyse a broad range of data types, with the ten Grand Challenges ICODA driver projects enabling access to datasets from over 70 countries. To date, this cohort of ten driver projects has generated 57 outputs, including publications, processes, dashboards, datasets for secondary analysis, and code.[15] Following review of the data research cycle used by each driver project, the lead researchers and ICODA team have worked together to identify common challenges in using secondary, health-relevant data in a global context to provide rapid insights and set out solutions they used that could be applied by others across other health challenges and data types.

Methods

The cohort of ten Grand Challenges ICODA driver projects all focused on major research questions related to the COVID-19 pandemic, but used different types of secondary data from a wide range of sources and contexts. As the projects were established at similar times, the ICODA team worked with the project leads to conduct a retrospective analysis of the challenges that each project faced in going through the relevant stages of the health data research cycle, and to collect project timeline data that clearly identified where common challenges and barriers existed across the cohort. The teams also worked together to identify the approaches and solutions that helped address these challenges, and which could be used by other researchers taking data science approaches to address a wide range of different health challenges.

Common process steps were selected for measurement in each of the projects, following the broad steps for the

	Main aims	Types of data and data source	Country	Key challenges	Key solutions
The PRIEST study for low- and middle-income countries (DP-PRIEST)	To ensure hospitals in low-income and middle-income countries are not overwhelmed during the COVID-19 pandemic by developing a risk assessment tool for clinicians to quickly decide whether a patient needs emergency care or can be safely sent home.	Existing data for 50 000 patients with suspected COVID-19 and who sought emergency care.	UK, South Africa, and Sudan	Obtaining additional approvals; linking and cleaning the datasets took longer than anticipated; projects involving teams distributed across the world require different methods for co-ordination than when all located in one place.	Team ultimately required a 3-month, no-cost extension to complete all planned analysis—a key lesson for use of any routine datasets. The ICODA initiative made expert curation support available and the use of natural language processing in data curation meant that the standardisation of the datasets could be achieved; the ICODA Workbench was extremely useful in providing a central repository of datasets to facilitate standardisation and analysis. There was a clear division of labour and responsibility between the different teams.
Addressing critical COVID-19 questions through research using linked population data (DP-ACCORD)	To understand COVID-19 evolution and impact, also on pregnancy and chronic diseases, by applying a data science approach to health data to study the clinical epidemiology and evolution of a new SARS-CoV-2 variant, which emerged in South Africa.	Anonymised COVID-19 health data from the government health department including >1 million tests and 60 000 hospital admissions in the Western Cape province of South Africa.	South Africa	The biggest challenge was dedicating person-time to the analyses when there were resurgences and competing service priorities. This challenge could be unique to the project being led by health-service staff. Outcome (COVID-19 relatedness of morbidity and mortality) and exposure ascertainment (previous infection) became increasingly challenging over time.	Where the team were able to automate the updating of analysis datasets, repeating analyses as the epidemic progressed became progressively easier. This enhancement is reflected in the severity analyses based on a standard case cohort, which was updated daily.
Effectiveness of COVID-19 vaccination in Brazil using mobile data (DP-EFFECT)	To quantify the real-world value of COVID-19 vaccines for protecting individuals from severe disease, and for protecting the entire population from being infected.	Data from the national vaccination programme, as well as deaths and cases at a municipality level and from 43 hospitals.	Brazil	Research teams in low-income and middle-income countries might not always have worked within what are considered international best practice and standards for data governance and data sharing; access and analysis of health data are restricted to researchers and government representatives, and not available for the health workers and communities to rapidly respond to the crises.	The project provided an opportunity to update the team on current best practices in data governance, privacy, and sharing guidelines, using an open science approach. This was important to maintain good collaborations with international and national networks, based on research reproducibility and transparency; the team also developed dashboard monitors that enabled the local health and research teams and the local community to easily access and visualise data from the cohort studies. Using a dissemination strategy plan, the team communicated the results of their research to local health practitioners and residents during meetings and conferences.
Routine assessment of infections, prevention, and control of SARS-COV-2 in unequal populations (DP-RASUP)	To study COVID-19 transmission issues in socially and economically unequal populations, accounting for human behaviour, non-pharmaceutical interventions, and vaccine strategies. By developing a user-friendly surveillance platform, the community could follow up its risk and jointly contribute to decrease cases and mortalities.	Five datasets were reviewed: de-identified surveillance data—daily time-series of cases and deaths of confirmed SARS-CoV-2 infections (n=2 005 200); de-identified information on gender, age, comorbidities, and infectious status (n=345 281); socioeconomic determinants data at summary level—classifies municipalities according to welfare benefits measured by income, literacy, and housing (n=5570); human mobility data at summary level—variables affecting human mobility, given by both Google trends and the historical average daily flux data throughout the country using road, air, and fluvial networks (n=65 638 [daily flux data]; n=754 095 [total state time-series]); stringency index at summary level—a metric that the team constructed that summarised the level of governmental measures enacted by local states (n=803).	Brazil	Processing the data for real-time analyses and providing results to the community in an optimal way; data access is still a challenge. The team collected a large amount of data to understand the COVID-19 pandemic in Brazil. However, statistical and mathematical modelling is challenging, due to the lack of studies that adequately account for different sub-populations; development of communication materials tailored for targeted populations with different backgrounds and priorities.	Invest time in building a systematic data pipeline tailored to the needs of the project; redesign aspects of the project to incorporate real-world variables influenced by population inequalities into the literature; initially, journalists were primary consumers of the team's results, aiming to disseminate scientific knowledge about the spread of SARS-CoV-2 in the country to the public. However, recognising the importance of understanding the priorities of the local community, the team started to work with local communicators in the favelas of Salvador.

(Table continues on next page)

	Main aims	Types of data and data source	Country	Key challenges	Key solutions
(Continued from previous page)					
Evaluating social inequalities and their effects on the COVID-19 pandemic in a low-income and middle-income country (DP-IDS-COVID19)	To measure the social disparities to understand the extent to which, in terms of socioeconomic factors, demographic factors, and access to health care, vulnerable people become unwell and die from COVID-19.	Administrative data collected routinely by government institutions and publicly available. These included: Brazilian Census 2010, National Register of Health Facilities, Influenza Case Notification System, National Total Population and Estimated Age Groups for 2020, and Brazilian Index of Deprivation database.	Brazil	The main challenge was in the construction of the COVID-19 social disparity index. This was designed to define indicators that were statistically correlated with each other and which provided evidence of social determinants for COVID-19 infection. This was presented as a publicly available and interactive dashboard and there were challenges in ensuring best user experience. The short project duration was also challenging.	The team used the regionalisation of health services categorisation used by Brazil's publicly funded health care system, Sistema Unico de Saude. The social disparity indicators that were identified were reviewed by a range of stakeholders and this helped ensure consistency between regions. The team also carried out usability tests for the online dashboard to enhance the user experience.
Disruptions in clinical outcomes and care among patients with chronic conditions: a four-country retrospective cohort (DP-PIH. CovCo)	To understand how the COVID-19 pandemic has affected care provision, care use, and health outcomes among chronic care patients—specifically those with HIV, cardiovascular disease, or diabetes.	A retrospective cohort study among patients with diabetes, HIV, or hypertension receiving care at Partners in Health-supported facilities. The dataset included 111 252 electronic medical records of chronic care patients.	Haiti, Malawi, Mexico, and Rwanda	Timeline for extracting electronic medical records; navigating complex electronic medical records datasets and in varying formats.	Frequent discussion with multiple-site level and cross-site level team members to assist in facilitating the data. Also creating a staggered analysis plan to focus first on HIV patients and then on cardiovascular and diabetes patients; programming rules were designed to assist with data cleaning, while also discussing metrics among site-level experts.
Characterising COVID-19 transmission chains for precision mitigation using epidemiological survey data (DP-CHAIN)	To reconstruct transmission chains between individuals in households and communities, and study COVID-19 transmission patterns from the reconstructed transmission chains.	Line-list datasets included confirmed cases of COVID-19 with detailed information enabling linkage to other cases, such as close contacts and simultaneous presence in the same location. Data were compiled from publicly available case reports, released by governments, or extracted from published studies. The dataset encompasses approximately 40 000 cases from four Asian countries and regions.	China, Taiwan, Hong Kong, and Singapore	The initial challenge lay in the data merging process. This involves integrating various datasets that adhere to different standards and organisational structures. The task required identifying common elements within these disparate datasets and successfully combining them; the information extraction process, which involved extracting specific details such as dates, named entities, and the relationships between these elements from the texts. The process is both time-consuming and resource intensive.	The first solution involved a dedicated team of 20 data collectors who monitored government announcements daily to gather reports. This systematic approach ensured consistent and up-to-date data collection, using a high-standard human coding system. This system involves two independent coders for initial data processing, followed by a resolver to address any discrepancies. This method enhances the accuracy and reliability of the data, using computer-aided coding. This involves using the data processed by human coders to train neural networks. The performance of these neural networks is on par with that of human coders, thereby combining the benefits of both manual expertise and automated efficiency.
Assessing the resilience of health systems during COVID-19 using routine data (DP-REHCORD)	To assess the magnitude of disruptions for non-COVID-19 essential health services during the COVID-19 pandemic in ten countries.	In each country, the team compiled administrative or routine health information system data on the number of health services provided from 2019 to 2021. In Ethiopia, Ghana, Haiti, Laos, Nepal, and South Africa, the data were extracted from health management information systems. In the other countries, the team used data from the Sistemas de Información del Ministero de Salud (Chile), Sistema de información del Instituto Mexicano del Seguro Social (Mexico), the National Health Insurance Service Health Facility Claims Database (South Korea), and the National Health Database of the Ministry of Public Health (43-folders dataset; Thailand).	Chile, Ethiopia, Ghana, Haiti, Laos, Mexico, Nepal, South Africa, South Korea, and Thailand	Data harmonisation: several indicators had different definitions across countries; data cleaning: missing values continue to be an important problem in health management information systems data in many countries; lack of master facility lists: in some countries, it was difficult to obtain an official count for the number of facilities that should be reporting every month.	Data harmonisation: multi-country codebooks with clear definitions highlighting differences by country were developed; data cleaning: a standardised data cleaning process was developed to exclude health facilities with sparse reporting. The data cleaning code was made available on GitHub for improved transparency and reproducibility; lack of master facility lists: to assess completeness, the team used the maximum number of facilities reporting in any given month as the estimated maximum number of facilities. In some countries, the team had to rely on aggregate analyses at district or provincial levels.

(Table continues on next page)

health data research cycle set out in figure 2. Some of these steps were time limited for the projects to complete, such as data curation or initial analysis. Other steps were not key to project completion but were mandatory in the ICODA framework and principles; for example, making data findable, accessible, interoperable, and reusable (FAIR) via publication of metadata and access request routes, so other researchers are able to use it in the future. The key process steps tracked are outlined below.

	Main aims	Types of data and data source	Country	Key challenges	Key solutions
(Continued from previous page)					
Data descriptor, reference coding, and characterisation of the systemic complications of critical care patients included in the ISARIC COVID-19 dataset (DP-ISARIC)	To identify and develop tools and strategies to enhance the extraction and comprehensive utilisation of data within the ISARIC COVID-19 dataset, with the specific aim of identifying risk factors associated with systemic complications and assessing their effect on clinical outcomes.	Clinical data from patients hospitalised with COVID-19 globally shared as a part of the ISARIC Clinical Characterisation Group collaboration. ISARIC has assembled the world's largest global database on COVID-19 clinical data with detailed individual patient data on 657 312 hospitalised individuals from 1297 institutions across 45 countries. This database includes data from more than 705 000 patients and 1500 centres worldwide.	>60 countries	Standardisation and mapping of the data encoded in open text fields of the dataset; due to the large dataset size, this was divided into 16 individual large tables that were hard to manipulate.	Team developed a computational code that identified the most frequently used texts and mapped them to standard codes with relevant clinical meaning. They then reviewed these codes manually to ensure they were accurate, capturing the original text reported in the dataset; the team designed a computational strategy that allowed them to identify each subject in each table to extract the individual data registered in each table. Then, they created a smaller dataset that allowed them to perform comprehensive statistical analyses. The code used to generate these datasets was registered in an open platform that other researchers could use to conduct their extractions and analyses once they had access to the dataset.
Using routine data to understand adverse pregnancy and neonatal outcomes associated with the COVID-19 pandemic in Kampala, Uganda (DP-IROC)	Incidence and risk factors for COVID-19 among pregnant and lactating women and their infants.	Quantitative data from the electronic medical records system.	Uganda	Team had a lot of unstructured data collected using free text fields in the variables that were required for the analysis dataset; working in the trusted research environment was challenging, due to limitations on data table editor and bandwidth.	The team, with support from MMS Holdings, completed additional data curation processes to produce an analysis-ready dataset; team held interactive sessions with the ICODA Workbench provider to inform expansion of workbench space and utility by using a Windows Virtual Machine.

ICODA=International COVID-19 Data Alliance. ISARIC=International Severe Acute Respiratory and Emerging Infection Consortium.PRIEST=Pandemic Respiratory Infection Emergency System Triage.

Table: **Grand Challenges ICODA driver projects**

Signing a data processing agreement

These agreements were between the data custodians (institutions or organisations sharing the data) and the provider of the ICODA Workbench, Aridhia Informatics (Glasgow, UK). The agreements covered the legal obligations and data governance responsibilities for adding data to the ICODA Workbench, and all ICODA driver projects were provided with a data processing agreement by Aridhia Informatics, along with support from the core programme team to aid completion.

Gaining researcher accreditation

This step involved the project teams completing the accreditation process for all researchers requiring access to the research environment, in line with the safe people principle (panel 1). Researchers' suitability to be given access to the ICODA Workbench was assessed against the following criteria: all requested information was provided; researcher is affiliated to a legitimate organisation conducting research and a bona fide researcher; and researcher has professional qualifications and experience to work with health data. This suitability was assessed by ICODA's research manager.

Preparing or curating the data

This step encompassed activities related to loading and preparing the data for analysis. Two of the projects received support from a commercial company specialising in data curation to expedite the process and address the challenges of harmonising datasets from multiple countries. This process included missing data and non-standardised records, especially those in free text format.

Sharing the metadata

This step involved the research teams publishing the details of the data (metadata) used for their research project. The FAIR principles[16,17] underpinned the ICODA initiative, and projects were able to access support to list metadata descriptions of all the data used—including associated data access requirements—on an accessible metadata catalogue, with digital object identifiers that could then be cited in associated publications.[18] The core ICODA team assisted project members to perform this step, which often was left until the latter part of the projects.

Initial analysis

This step included the time to complete the first iteration of analysis of the research data. Several analysis tools

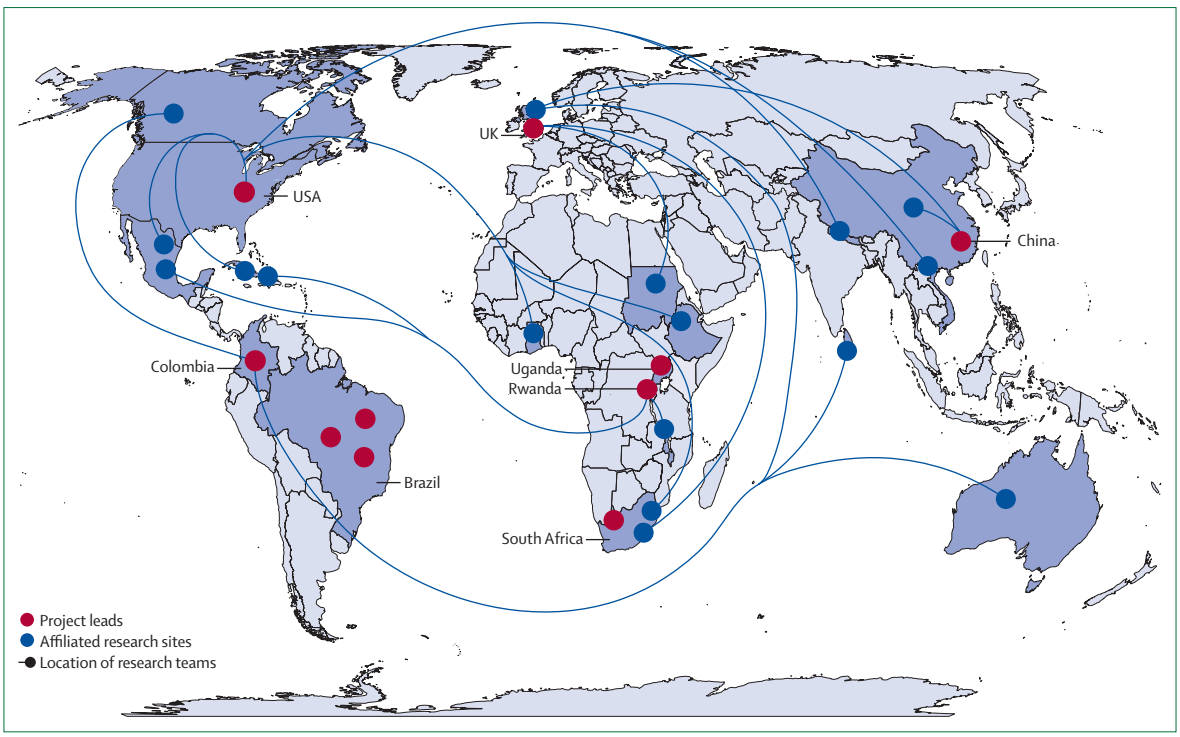


Figure 1: **Global scope of ten Grand Challenges ICODA driver projects**
The countries shaded dark blue indicate the geographical scope of the ICODA initiative.

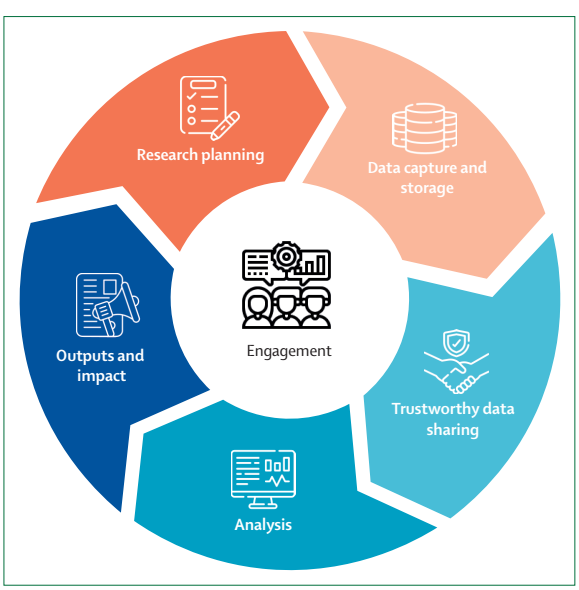


Figure 2: **Health data research project lifecycle**

were provided for use in the ICODA Workbench; in addition, bespoke tools were provisioned where required.

Sharing outputs

Each project addressed major research questions relating to the COVID-19 pandemic and aimed to generate rapid insights within 12 months. This step included activities related to initial publications for each research project as well as other outputs such as dashboards, community engagement materials, videos, and documentaries.

Engagement with and involvement of local communities, health practitioners, and policy makers

From the initial project design and through each of these process steps, the driver project research teams engaged with a range of stakeholders, including community members, health practitioners, and policy makers, to ensure research questions were relevant, the value of the research was understood, and outputs benefited the health of communities. Several projects invested substantial time and resources in engaging with local communities including: participation in a vaccine programme alongside a local non-governmental organisation; working with community leaders, community groups, and young influencers to communicate the purpose and value of the research; and engagement with local media.

A quantitative and qualitative analysis of the common process steps and the length of time required for each step was then carried out for the ten Grand Challenges ICODA driver projects, to identify common barriers and bottlenecks as well as possible solutions. The length of time taken for each process step was self-reported, in response to requests for this information. This information was verified through cross-referencing relevant correspondence and system-generated timestamps

where available. Although this information offered insights into the effort required at each process step, there are limitations to the data obtained, which can introduce bias and variability. These limitations include differences in the types of data, the analytical methods and the research approach being used by each driver project, as well as different interpretations of definitions of each process step, and the fact that the information was collected retrospectively. As can be observed, the time for each process step varied considerably between each project team with process step duration being shown in figure 3 (appendix).

Qualitative analysis was supported through the quarterly monitoring reports and final report provided by each driver project team, which specifically requested feedback on any lessons learned. Online meetings during the funding period and a mid-project convening of the cohort (focused on community and stakeholder engagement) also provided opportunities for feedback from the research teams.

Results: challenges and solutions for international health data access and sharing

Challenges within the health data research cycle

Analysis of the time taken for the ten Grand Challenges ICODA driver projects to undertake each process step in the health data research cycle is set out in figure 3. This analysis highlighted some common challenges and barriers, and suggested possible solutions in relation to putting data processing agreements in place, researcher accreditation, data curation and preparation, ensuring metadata was prepared and published appropriately following FAIR principles, and undertaking analysis of secondary data. These and other challenges that research teams using data science approaches to address health challenges might have will be explored further.

Identifying data sources

A secondary data use project usually starts with identifying existing data to answer the research question. This process often involves a literature search and scanning other sources where available datasets are listed, such as repositories and data catalogues. Despite improvements in digital cataloguing, this stage is still challenging due to a lack of interconnectedness in the data sharing ecosystem and inconsistent use of metadata standards around the world. The Grand Challenges ICODA driver projects required data to have been already identified at the application stage, so data source identification time was not included within the project timelines.

Data processing agreements

A median of 1·5 weeks of project time was taken up with finalising data sharing agreements. This figure was calculated using data reported by seven of the ten project teams. Research contracts are recognised to be time-consuming, as multiple departments get drawn in to reviewing and negotiating terms. Contracts relating to data reuse are particularly complex due to additional regulatory requirements, such as data protection obligations. One solution to reduce the time taken for this stage is to standardise data sharing agreements as much as possible, as was done for the ICODA programme. Standardisation would ideally use an internationally agreed set of minimum criteria that could be used by data custodians and data users around the world.

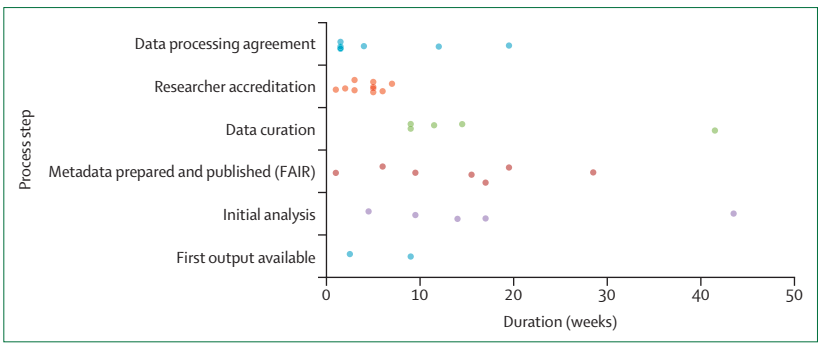


Figure 3: **Strip plot showing the time taken in weeks across the ten driver projects to complete each process step**
FAIR=findable, accessible, interoperable, and reusable.

See **Online** for appendix

Contract negotiation difficulties stemmed from both the complexity of agreements and models to enable data sharing. Each institution was responsible for uploading their data into the TRE, and driver project teams were encouraged to make their data FAIR. Aridhia Informatics, the TRE provider, provided their standard data processing agreement, which was used for projects to access the TRE. Projects were responsible for their own data uploaded to the TRE.

Each institution was recognised as a data controller of the data they contributed, enabling the use of their data within the TRE; however, this limited the ability to share data between projects.

Researcher accreditation

Researcher accreditation was a quick process stage for researchers to complete. It was automated via the ICODA Gateway; data reported by all ten of the project teams indicated a median of 5 weeks for each project. The process also allowed for additional researchers to be brought into project teams quickly where required.

Data curation and preparation

Most datasets need to be curated before analysis can begin, which is typically a time-consuming step. The median time spent curating data was 11·5 weeks, but there were considerable outliers. The median time taken was calculated using data reported by five of the ten driver project teams.

A natural language processing approach was applied in the case of the DP-PRIEST study to address the challenges of converting free text to clinical codes. The

Panel 2: **Challenges and solutions from the ten Grand Challenges International COVID-19 Data Alliance (ICODA) driver projects**

Identifying available data

Challenges

It is difficult to identify available data globally, with limited connections between repositories and data catalogues, and no internationally agreed minimum metadata standards.

Published articles regularly have limited or no data sharing statements included.

Solutions

Researchers should always list metadata on a publicly accessible repository and include clear data sharing statements with as much detail as possible about access request requirements. By the end of their grants, nine of ten Grand Challenges ICODA projects had listed their metadata on the ICODA Gateway in a standard format to enable other researchers to understand the contents of the datasets and request data access.[6]

Each metadata description includes a digital object identifier which can be cited in relevant articles to ensure appropriate attribution.

All articles published by the teams should include data availability information, along with a digital object identifier.

Data sharing agreements

Challenges

Getting contracts signed is a protracted process and terms are often debated.

Solutions

Encourage use of standard templates and share these with relevant contract departments as early as possible.

Simplified language is key, with agreements needing to be approachable and clear while serving the required purpose.

When carrying out research in an international context, it is important to identify and share best practice in data governance and privacy and share guidelines with teams, especially when taking an open science approach.

Accrediting researchers

Challenges

Protecting the privacy of participants and preventing misuse of patient data are critical when using health data for research and it is key that trustworthy and transparent systems are built into all activities undertaken by researchers.

Solutions

ICODA made use of the Five Safes framework first developed by the UK's Office for National Statistics and adopted by Health Data Research (HDR) UK.[10]

ICODA researchers were assessed against Safe People criteria and this is an approach we would recommend for adoption by others as a proportionate review process for data access to ensure data is only accessed by trained and accredited researchers who are trusted to use it appropriately.[11]

Data curation and challenges preparing data

Challenges

Lack of data standardisation, even within single health datasets, means that combining data from multiple sources can be extremely difficult and time-consuming to convert and transform.

Solutions

For ICODA projects, the initial aim of working to a standardised data dictionary across the spectrum of projects was realised to be too ambitious and was therefore refocused on preparing dictionaries for individual projects, rather than across the initiative.

Data curation tasks were aided by partnerships with expert teams, initially with Aridhia Informatics and Cytel for the first ICODA driver project, and then with MMS Holdings for the ten Grand Challenges ICODA driver projects. These projects had mixed data curation needs and being able to ask questions of the statistical expert group and expert curation partners helped formulate the project teams' thinking.

Having professional partnerships available to all projects earlier would have facilitated progress of research and been a worthwhile investment.

Sharing tools for transforming data, particularly in open-source format, can help the secondary data use community. ICODA supported shared tools within the trusted research environment, and shared tips and approaches for data curation through webinars and workshops.

Realistic costs for data curation and transformation need to be included upfront and planned for in grant budgets, along with sufficient time for this activity built into the project plan.

Making metadata available

Solutions

HDR UK's existing Innovation Gateway was repurposed to support the ICODA initiative, creating a new Gateway instance with associated branding.[6] The ICODA Gateway enabled metadata to be made visible, while also implementing key processes, such as researcher accreditation and data access requests. Researchers entered metadata into the metadata catalogue provided within the trusted research environment which was then federated to the ICODA Gateway, making metadata publicly visible.

This reuse of HDR UK's existing software assets accelerated time to delivery through customising, rather than coding from scratch, which was cost-saving and time-saving. Piloting metadata federation with the ICODA trusted research environment partner resulted in functionality that has subsequently been implemented more broadly within HDR's Innovation Gateway and is being rolled out across the HDR UK ecosystem.

(Continues on next page)

(Panel 2 continued from previous page)

Ensure training is available for teams to upload metadata to catalogues and project time is planned in for curating the metadata or data dictionary itself. Many teams left this step until the later stages of their projects and data could have potentially been reused within the cohort of projects and more broadly, had this been done earlier.

Enabling data analysis

Challenges

Still difficult for sensitive data or multiple large datasets or data spanning country borders.

Solutions

Researchers in the ICODA programme were provided with access to the COVID-19 Workbench, a cloud-hosted trusted research environment delivered by our partner, Aridhia Informatics. This provided researchers with a secure space to perform collaborative research, with controlled access, scalable compute power, tooling, assistance with data provision, hosting, and support.

Use of a turnkey, cloud-hosted trusted research environment jumpstarted the initiative and is a route to be considered for future projects. It was key in enabling collaboration, saved the creation of multiple instances hosting the same data (important when working on large datasets across geographical locations), provided data custodians with a high level of confidence in security of data, and proved accessible from low bandwidth settings, providing high end computational power to research teams where required.

Effective community and stakeholder engagement

Challenges

Teams often had limited experience of community and patient engagement in shaping research projects, and pandemic restrictions made engagement even more challenging.

The limited time scale (12 months) of the projects also proved challenging.

Solutions

The ICODA team ran workshops and question and answer sessions on stakeholder and community involvement and engagement for long-listed research teams to further develop their detailed plans and, working with expert groups, convened a community, public, and patient review panel for the Grand Challenges ICODA open funding call.

A halfway convening workshop was organised for all ten Grand Challenges ICODA driver project teams, which focused on their community, public and patient involvement and engagement plans, progress, and challenges, enabling knowledge sharing across the cohort. With guidance from ICODA's ethics advisory council, an ethics and governance framework was developed for use by all project teams.[18]

Engaging stakeholders in setting research questions early and the communication of results proved valuable. This included establishing a dissemination plan with local stakeholders based on active listening with rapid communication, as used by DP-EFFECT.[19] Social media activities and partnership with local influencers, including primary care workers and community leaders, were cited as important.

Creating different communication approaches for different audiences, including videos and webinars, was valuable, and the DP-IDS-COVID-19 team highlighted the importance of avoiding use of technical terms.[20] The best tool to be used depends on the objective of engagement with participants.

Budget for engagement activities and ensure that there are members in the team with the appropriate skills to support this work.

Partner with local civil society organisations and the public and private sectors to ensure local needs are embedded in the research.

Engaging policy makers in using research outputs

Solutions

Plan for this element from project initiation and engage early to ensure outputs are in a useful and consumable format for policy makers. Several teams produced dashboards for use by policy makers or communities, including DP-IROC[21] and DP-ACCORD.[22]

Tailor communications and create and circulate regular policy briefings.

Set up and hold regular briefing meetings to highlight the importance and relevance of the research to policy development, change, or implementation.

DP-CHAIN project also used natural language processing to curate their data and the natural language processing algorithm was published on iScience.[19] The DP-REHCORD team highlighted the importance of setting up a clear and standardised codebook when preparing datasets for analysis; for example, it was suggested that codebooks should include variable names, labels, and any skip patterns.[20] This method was found to be particularly important if working across multiple datasets, in large teams, or across multiple countries.

Another recommendation was to adopt a version control system for statistical code such as GitHub, to allow for collaborative and simultaneous work across large teams to ensure that code is not lost or overwritten. Making this code publicly available through such platforms also improves reproducibility and transparency. For studies in countries with disaggregated data, teams were advised to implement simple and standardised procedures for data cleansing. DP-CHAIN's codebook and data are published.[21] The DP-ISARIC team[22] applied a uniform data model to standardise the structures and ontologies in the ISARIC dataset to a harmonised format. All data were standardised to the Clinical Data Interchange Standards

Consortium Study Data Tabulation Model to facilitate pooled analyses.

Making metadata available

Median time spent making metadata available was 15·5 weeks, from data reported by seven of the ten project teams. This duration, although lengthy, reflects not only the time taken to upload the metadata to an online catalogue, but also the work involved in preparing the metadata itself. Organisation of variables, their names, descriptions, and valid ranges ended up being a substantial piece of work for many teams.

For more on **ICODA news and events** see https://icoda-research.org/news-and-events/

Initial analysis

Driver project teams experienced first-hand the challenges of using diverse health data collected from multiple sources and trying to combine and analyse them. One of the main difficulties encountered in combining data was the lack of data standardisation, quality, and structure, which reduced interoperability. Databases often had missing information and unstructured entries, such as free text information, which made analysis difficult.

The use of additional statistical and data science expertise to help execute data analysis was noted by a number of teams; the DP-RASUP team highlighted the importance of not underestimating the time needed for processing data, especially large and complex datasets.[23] Median time taken for the initial analysis stage was 14 weeks, using data reported by five of the ten project teams.

An example of how these challenges were overcome comes from the DP-CHAIN project, which reconstructed transmission pairs using epidemiological survey data published by governments around the world.[19] Different countries adopt different standards when tracing close contacts and report their findings in different ways. DP-CHAIN standardised global contact tracing data by categorising them into two types: individual contacts and contact clusters. The team further developed an algorithm to infer transmission pairs from contact networks. Epidemiological characteristics, such as the basic reproduction number (R_0) and dispersion, could then be calculated from the transmission networks and compared across countries. R_0 is the average number of secondary infections generated by a single infectious individual in a population where all members are susceptible to the infection, with higher values indicating an increase in infection in the population.[24] The dispersion parameter (k) quantifies the variation in the number of secondary infections caused by infected individuals, with lower values of k indicating a greater likelihood of superspreading events.[25] Taking an innovative approach, the DP-RASUP team documented their methods for data processing, data analysis, data visualisation, mathematical modelling, and statistical modelling in a series of YouTube videos, which are freely available for other researchers to view and learn from.[23]

For the **digital hub** see https://globalhealthdatascience.tghn.org/

For the **policies and processes** see https://icoda-research.org/research/publications/#genericgovernanceprocessesforreus

ICODA established several support mechanisms to enable teams to overcome analysis challenges. For example, a team of statistical experts were identified who provided advice and input to the research teams where needed; they commented on statistical analysis plans and developing and sharing analysis tools with them as the projects progressed. Furthermore, tools, code, and curation advice were shared between ICODA project team members within the analysis environment and more broadly through data science webinars and workshops, many of which are available to watch online on the ICODA website. A full summary of the challenges, solutions, and lessons learned by the ten Grand Challenges ICODA driver projects is presented in panel 2.

Outputs and impact

Across the ten ICODA driver projects, a wide range of project outputs were planned and delivered including: results manuscripts, code, methods papers, dashboards, community and stakeholder engagement materials and tools, videos, and documentaries. Despite the challenges outlined in this Health Policy paper, all teams were successful in delivering rapid insights and outputs, with some projects beginning to share findings as early as 6–7 months into the project. Most teams published their findings between 9 months and 15 months after the projects started, with all publications, code, and metadata being made open and accessible to other researchers.

Innovative outputs included those from the DP-PRIEST team who produced a validated triage tool for use by clinicians in low-income and middle-income settings for assessing whether patients should be admitted with COVID-19 to intensive care units.[26] Other teams have documented their methods, community engagement experiences and approaches, as well as tools for use by policy makers and health service leads. The DP-IDS-COVID19 team developed an index to measure social inequalities during the pandemic in Brazil, which was used by a council of representatives of state health managers to identify people vulnerable to COVID-19 and guide the planning of interventions. These outputs are shared more widely on the Global Health Data Science digital hub, to which the ICODA teams have contributed.

The ICODA initiative has sought to maximise research impact through making a range of outputs openly available, including transformational code, dataset metadata (on HDR's Innovation Gateway), community engagement materials, and governance policies and processes. These policies and processes are available on the ICODA website and have been genericised for wider reuse.

A less tangible but equally important outcome of the initiative has been that a global health data science community of practice has been established and continues to be active. The ten Grand Challenges ICODA

projects are now firmly embedded in the wider Grand Challenges community and their research teams continue to engage as part of this wider data science community of practice through the Global Health Data Science digital hub and HDR UK's Global programme.

Further perspectives and wider lessons learned

Engagement with local communities, health practitioners, and policy makers

Community and stakeholder engagement at all stages of the data reuse project cycle underpins relevant and quality research. It was a key element of the ten Grand Challenges ICODA driver projects, having been built into the design of the global funding call, and the subsequent support for the driver projects to deliver benefits to patients and better health outcomes for all. Levels and types of engagement varied across projects and involved direct engagement with local communities to raise awareness of the research and the potential positive effect on community health priorities through using data science-enabled insights to influence health policy and practice. Direct engagement with policy makers and practitioners also took place to shape research projects and ensure insights and outputs were taken up.

Teams had several challenges associated with engagement activities, but also found innovative solutions. The DP-EFFECT team used mobile app-based technologies to provide rapid access to free COVID-19 testing during the pandemic, showing the high potential of these e-health technologies in improving the access of vulnerable populations to health-care services.[27] Since completing the Grand Challenges ICODA research, the DP-EFFECT team has been developing a community engagement toolkit based on their stakeholder engagement experiences, from which other researchers can benefit. The DP-PRIEST team set up a public patient involvement and engagement group in the Western Cape, South Africa, including eight community members affected by COVID-19 (infected themselves or an immediate family member was infected or hospitalised). The group were kept informed about the study, and then given the opportunity to provide feedback. The feedback provided was particularly useful for gaining a public perspective on the use of anonymised routinely collected health-care data. Engagement was focused on raising awareness of the research and the findings and their potential to inform health policy.

The DP-IDS-COVID19 team invested substantial time engaging with a range of stakeholders and published their engagement experience in a journal article.[28] They describe how community members and policy makers made contributions to existing research through informing a new layer of information in the interactive social disparities index developed by the team, as well as improvements to the interactive index panel itself. Eight representatives of community groups and 29 policy makers participated in engagement activities during the project, more than 500 people engaged in open webinars about the project, and over 140 news items about this study were published in national and international media.

Research impact and uptake rely on the involvement of all key stakeholders through the research lifecycle. Challenges faced by the ICODA driver project teams included the fact that ministries of health were occupied with managing the COVID-19 burden and fast-moving policy development, making engagement difficult. Despite these challenges, the DP-IDS-COVID19 and DP-IROC teams developed dashboards that were accessible by ministries to improve targeting of COVID-19 interventions.[28,29] The DP-PRIEST research team engaged with a range of clinical academics from the South African and Sudanese health-care settings as well as local networks to discuss findings and the acceptability of triage tools developed.

Conducting research during a pandemic

Conducting any research in a pandemic is challenging and secondary health data use projects were no exception. COVID-19 restrictions limited in-person meetings and networking, something particularly important in building trust and productive teamwork across multiple locations, and the involvement of community and other stakeholders. Maintaining up-to-date data in a fast-changing context was difficult, as was implementing dynamic data collection with fast clinical and epidemiological variations.

A further challenge in conducting research for studies led by health service staff, was dedicating person-time to the analyses when there were resurgences of COVID-19 and competing service priorities. To address this, the DP-ACCORD team were able to automate the updating of analysis datasets, so that repeating analyses as the epidemic progressed became progressively easier.[30]

Conclusion

The ICODA programme itself has now completed, and achievements and outputs from the initiative and its driver projects have been significant and varied. To date, there have been 38 publications from its full cohort of 12 driver projects, with more papers submitted. All achievements and outputs are available to other researchers through open access publication and are indexed on the ICODA website.[15]

This Health Policy paper summarises the common challenges and potential solutions to accelerate the health data research cycle, based on the research process of a diverse range of ten research driver projects under the ICODA programme, and highlights where data infrastructure and governance could be improved to better enable secondary data use studies. The approaches, tools, and outputs from these driver projects are openly available and can be used by researchers for similar

studies using secondary data to address a wider range of health challenges, and the implementation of some of these solutions early on could help to accelerate the generation of insights and outputs.

Contributors

SB led on writing of the manuscript and GSH conceived and wrote the first draft. The principal investigators of the Grand Challenges ICODA driver projects (CA, FAB, MYI, KlD, XFL, JCM, JFO, and LFR) and team members (LH, JO, MB, and AF) contributed to the writing, reviewing, and editing of the manuscript. From the ICODA core team, MR led on data analysis from the studies and ADM, NP, MR, EM, and AW contributed to writing, reviewing, and editing the manuscript. EF contributed to the review, editing, and writing of the manuscript.

Declaration of interests

All coauthors (except EF) received funding from the Bill & Melinda Gates Foundation and Minderoo Foundation for the ICODA initiative. KLD received grant funding for the main project from the European and Developing Countries Clinical Trials Partnership (EDCTP) 2 programme supported by the European Union (RIA2020EF-2926 periCOVID Africa). LFR received grants from Pfizer and Merck Sharp & Dohme; consulting fees from Merck Sharp & Dohme, Pfizer, and GSK; payment for lectures and presentations from Merck Sharp & Dohme and GSK; and payment for expert testimony from Merck Sharp & Dohme, Pfizer, and GSK. MYI received grant funding from the Ministry of Health (decentralised executive term, process number 25000200517/2019-46), National Institute for Health and Care Research (NIHR), and the Wellcome Trust; and payment to participate in the International Population Data Linkage Network Conference from the NIHR Global Health Research Program Award. MB received grants from the London School of Economics, The Rockefeller Foundation, Global Challenges Research Fund Global Multimorbidity, and Google; has received consultancy fees from the Singapore Institute of Management; and has a patent from the Institute of Electrical and Electronics Engineers Standards Association. AF has received grants from the Ministry of Health, NIHR, and the Wellcome Trust. GSH received funding from the Gates Foundation contract for contract work unrelated to the contents of the manuscript; consultancy fees from Vivli clinical data sharing platform and ClinicalStudyDataRequest.com clinical data sharing platform; and fees from the National Institute on Ageing to present at their data sharing workshop. All other authors declare no competing interests.

Acknowledgments

We acknowledge and give special thanks for Joseph Ouma's contribution to the writing of this manuscript, the delivery of a webinar, and to the Grand Challenges ICODA initiative. Very sadly, Joseph passed away on Sept 2, 2023, a profound loss to his family and colleagues. We dedicate this paper to the life and work of Joseph Ouma and extend our deepest condolences to Joseph's family and loved ones. We also thank and acknowledge the significant contributions of the patients and public whose data informed the research projects; all members of the ten Grand Challenges ICODA research driver project teams who have participated in this initiative; and the data contributors who willingly shared their datasets for reuse and for the benefit of patients and the wider public. We also thank Steve Kern for his extensive guidance and commitment to the initiative; Kim Carter for his engagement and unwavering support; Amel Ghouila for her advice and inputs to the design and delivery of the Grand Challenges ICODA open funding call; Carl Marincowitz for his engagement in the ICODA initiative and participation in a webinar on Trusted Research Environments; Agklinta Kiosia for her review of the paper; and our technical delivery partners, Aridhia Informatics, Certara, Cytel, MMS Holdings, PA Consulting, Preva Group, and SAIL Databank for their contributions. The ICODA initiative was shaped and supported by three key governance bodies: the Scientific and Strategic Advisory Council, the Ethics Advisory Council, and the Executive Leadership Team. We thank everyone involved for so generously giving their time and expertise to guide the initiative to success. This work was financially supported by ICODA, an initiative funded by the Gates Foundation (INV-017293), the Minderoo Foundation, supported by Microsoft's AI for Good Research Laboratory, and convened by Health Data Research UK. Aridhia Informatics was funded by the Gates Foundation (INV-021793) to provision the Workbench for the driver projects. The views and opinions of the authors expressed herein do not necessarily reflect those of EDCTP. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

References

1 Science Academies of the Group of Seven (G7). Data for international health emergencies: governance, operations and skills. March 31, 2021. https://royalsociety.org/-/media/about-us/international/g-science-statements/G7-data-for-international-health-emergencies-31-03-2021.pdf (accessed July 25, 2023).
2 Moorthy V, Henao Restrepo AM, Preziosi MP, Swaminathan S. Data sharing for novel coronavirus (COVID-19). *Bull World Health Organ* 2020; **98:** 150.
3 Wellcome. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Jan 31, 2020. https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak (accessed July 25, 2023).
4 Baynes G, Hahnel M. Research practices in the wake of COVID-19: busting open the myths around open data. https://www.springernature.com/gp/advancing-discovery/springboard/blog/blogposts-open-research/research-practices-in-the-wake-of-covid/18256280 (accessed July 25, 2023).
5 Näher AF, Vorisek CN, Klopfenstein SAI, et al. Secondary data for global health digitalisation. *Lancet Digit Health* 2023; **5:** e93–101.
6 Benchimol EI, Smeeth L, Guttmann A, et al. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med* 2015; **12:** e1001885.
7 Bernier A, Raven-Adams M, Zaccagnini D, Knoppers BM. Recording the ethical provenance of data and automating data stewardship. *Big Data Soc* 2023; **10:** 20539517231163174.
8 Sigfrid L, Maskell K, Bannister PG, et al. Addressing challenges for clinical research responses to emerging epidemics and pandemics: a scoping review. *BMC Med* 2020; **18:** 190.
9 Desai T, Ritchie F, Welpton R. Five Safes: designing data access for research. 2016. https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf (accessed July 25, 2023).
10 International COVID-19 Data Alliance. Research review process. October, 2022. https://doi.org/10.57775/mhbw-2v62 (accessed Aug 7, 2023).
11 International COVID-19 Data Alliance. Generic output research review process for research projects. October, 2022. https://doi.org/10.57775/6jph-dt22 (accessed Aug 7, 2023).
12 Health Data Research (UK) Innovation Gateway. International COVID-19 data alliance datasets. June, 2023. https://web.www.healthdatagateway.org/collection/29513503379534136 (accessed Aug 7, 2023).
13 Dron L, Kalatharan V, Gupta A, et al. Data capture and sharing in the COVID-19 pandemic: a cause for concern. *Lancet Digit Health* 2022; **4:** e748–56.
14 Calvert C, Brockway MM, Zoega H, et al. Changes in preterm birth and stillbirth during COVID-19 lockdowns in 26 countries. *Nat Hum Behav* 2023; **7:** 529–44.
15 International COVID-19 Data Alliance. Publications and outputs. 2022. https://icoda-research.org/research/publications (accessed Jan 9, 2024).
16 Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016; **3:** 160018.
17 Maxwell L, Shreedhar P, Dauga D, et al. FAIR, ethical, and coordinated data sharing for COVID-19 response: a scoping review and cross-sectional survey of COVID-19 data sharing platforms and registries. *Lancet Digit Health* 2023; **5:** e712–36.
18 International COVID-19 Data Alliance. ICODA metadata instructions. 2024. https://icoda-research.org/wp-content/uploads/2024/01/ICODA-metadata-instructions.pdf (accessed Jan 26, 2024).
19 Wang Z, Liu XF, Du Z, et al. Epidemiologic information discovery from open-access COVID-19 case reports via pretrained language model. *iScience* 2022; **25:** 105079.
20 Arsenault C, Gage A, Kim MK, et al. COVID-19 and resilience of healthcare systems in ten countries. *Nat Med* 2022; **28:** 1314–24.

21 Liu XF, Xu XK, Wu Y. Mobility, exposure, and epidemiological timelines of COVID-19 infections in China outside Hubei province. *Sci Data* 2021; **8:** 54.

22 Garcia-Gallo E, Merson L, Kennon K, et al. ISARIC-COVID-19 dataset: a prospective, standardized, global dataset of patients hospitalized with COVID-19. *Sci Data* 2022; **9:** 454.

23 Pereira FAC, Filho FMHS, de Azevedo AR, et al. Profile of COVID-19 in Brazil-risk factors and socioeconomic vulnerability associated with disease outcome: retrospective analysis of population-based registers. *BMJ Glob Health* 2022; **7:** e009489.

24 Boonpatcharanon S, Heffernan JM, Jankowski H. Estimating the basic reproduction number at the beginning of an outbreak. *PLoS One* 2022; **17:** e0269306.

25 Wegehaupt O, Endo A, Vassall A. Superspreading, overdispersion and their implications in the SARS-CoV-2 (COVID-19) pandemic: a systematic review and meta-analysis of the literature. *BMC Public Health* 2023; **23:** 1003.

26 Marincowitz C, Sbaffi L, Hodkinson P, et al. Prognostic accuracy of triage tools for adults with suspected COVID-19 in a middle-income setting. *Emerg Med J* 2022; **39:** A976–77.

27 Ranzani OT, Silva AAB, Peres IT, et al. Vaccine effectiveness of ChAdOx1 nCoV-19 against COVID-19 in a socially vulnerable community in Rio de Janeiro, Brazil: a test-negative design study. *Clin Microbiol Infect* 2022; **28:** 736.e1–4.

28 dos Anjos Fonseca A, Pimenta DM, de Almeida MRS, Lima RT, Barreto ML, Ichihara MYT. Public involvement & engagement in health inequalities research on COVID-19 pandemic: a case study of CIDACS/FIOCRUZ BAHIA. *Int J Popul Data Sci* 2023; **5:** 2133.

29 International COVID-19 Data Alliance. Incidence and risk factors for COVID-19 amongst pregnant and lactating women and their infants in Uganda (DP-IROC). 2023. https://icoda-research.org/project/dp-iroc/ (accessed Aug 7, 2023).

30 Davies MA, Morden E, Rousseau P, et al. Outcomes of laboratory-confirmed SARS-CoV-2 infection during resurgence driven by omicron lineages BA.4 and BA.5 compared with previous waves in the Western Cape Province, South Africa. *Int J Infect Dis* 2023; **127:** 63–68.

Proposta de classificação dos diferentes tipos de estudos epidemiológicos descritivos

doi: 10.1590/S1679-49742021000100026

Proposal for classifying the different types of descriptive epidemiological studies

Propuesta de clasificación de los diferentes tipos de estudios epidemiológicos descriptivos

Edgar Merchán-Hamann[1] – orcid.org/0000-0001-6775-9466
Pedro Luiz Tauil[2] – orcid.org/0000-0002-9333-3327

[1]Faculdade de Ciências da Saúde, Universidade de Brasília, Brasília, DF, Brasil
[2]Faculdade de Medicina, Universidade de Brasília, Brasília, DF, Brasil

Resumo

A categoria dos estudos epidemiológicos descritivos é tema relevante, uma vez que existem inconsistências na literatura quanto a sua nomenclatura e classificação. Foram revistos livros de textos acadêmicos de epidemiologia, 19 estrangeiros e seis nacionais, sendo o critério principal tê-los disponíveis para revisão detalhada dos capítulos de epidemiologia descritiva e tipos de estudo. Em 11 livros, os autores dão prioridade aos estudos analíticos. Doze textos estrangeiros e dois brasileiros incluem estudos descritivos, apesar de a maioria não explicitar uma categoria específica com esse nome. Propõe-se uma classificação com base nas respostas a questões norteadoras de pesquisa, incluindo os seguintes tipos de estudos: relato de caso, série de casos, coorte clínica, estudo de prevalência, estudo de incidência (coorte) e estudo ecológico descritivo. Discutem-se as potencialidades do seu uso, a implementação de novos métodos de análise e sua relevância na vigilância à saúde.

Palavras-chave: Estudos Epidemiológicos; Estudos Descritivos; Classificação.

Endereço para correspondência:
Edgar Merchán-Hamann – Universidade de Brasília, Departamento de Saúde Coletiva, *Campus* Universitário Darcy Ribeiro, Brasília, DF, Brasil. CEP: 70910-900
E-mail: hamann@unb.br

Introdução

Segundo o *Dicionário de Epidemiologia* de M. Porta,[1] epidemiologia é o estudo da ocorrência e distribuição de eventos relacionados à saúde em populações específicas, incluindo o estudo dos fatores determinantes que influenciam tais eventos, e a aplicação desse conhecimento para controlar os problemas de saúde. O estudo da ocorrência e da distribuição de eventos constitui o objeto dos estudos epidemiológicos descritivos.

> *O estudo da ocorrência e da distribuição de eventos constitui o objeto dos estudos epidemiológicos descritivos.*

Nesses estudos, pesquisadores não buscam verificar a existência ou não de associação entre variáveis de exposição e de efeito. Tradicionalmente, define-se como uma característica primordial o fato de não terem um grupo de comparação, comumente chamado de "grupo controle". A determinação da frequência dos eventos e da sua distribuição, segundo as características das pessoas acometidas ou que relatam um determinado antecedente, localização espacial e temporal, permite identificar coletivos populacionais, áreas geográficas e épocas de risco (incidência) ou de maior presença (prevalência) do agravo. Permitem ainda formular hipóteses a respeito dos fatores responsáveis por sua frequência e distribuição. Tais hipóteses podem ser ulteriormente testadas mediante estudos epidemiológicos analíticos.

A finalidade deste artigo é caracterizar os diferentes tipos de estudos descritivos existentes, propor uma classificação e contribuir para a avaliação apropriada das suas potencialidades e limitações com relação aos objetivos pretendidos.

Antecedentes da caracterização dos estudos descritivos

Foram revistos livros de texto acadêmicos de epidemiologia da literatura internacional e nacional; o critério principal foi tê-los disponíveis para revisão detalhada dos capítulos de epidemiologia descritiva e tipos de estudo. Ao todo, o material que fundamenta a presente proposta corresponde a 25 livros de texto, sendo 19 estrangeiros e seis brasileiros. Tal material foi produzido por 27 autores ou grupos de autores. Em dois livros de texto, há dois capítulos de autores diferentes que tratam do tema.

A questão da nomenclatura dos estudos epidemiológicos descritivos começa pelo próprio reconhecimento de uma categoria que inclua tais estudos. Muitos autores simplesmente desconsideram ou não mencionam qualquer tipo de estudo epidemiológico que não seja analítico, isto é, que não tenha grupos para comparação e que não teste hipóteses.

Na Figura 1, são mostrados oito livros de texto estrangeiros, um com dois capítulos de autores diferentes, nos quais não há detalhamento dos estudos descritivos, porque se outorga absoluta prioridade aos estudos analíticos. Parte-se de uma das primeiras obras acadêmicas da área, o texto de MacMahon e Plugh (1970),[2] indo até uma das mais completas em sua reflexão teórica e aprofundamento metodológico, o texto de Rothman e colaboradores (2008).[3] Nos mencionados textos, apenas se apresenta a epidemiologia descritiva e diferenciam-se as medidas de incidência e prevalência.

Na Figura 2, apresentam-se 12 livros de texto em que há algum detalhamento de determinados estudos epidemiológicos descritivos, apresentados em seções ou capítulos específicos. Muitas vezes, tais capítulos não são intitulados como epidemiologia descritiva nem como tipos de estudos descritivos. Por exemplo, Gordis (2014)[4] menciona estudos descritivos nos capítulos "História natural da doença e prognóstico" e "Avaliação de medidas preventivas e terapêuticas". Em geral, os autores reconhecem a diferença entre estudos descritivos e analíticos, conferindo maior relevância ao segundo tipo. Na pesquisa de campo aplicada, apesar da valorização dada à epidemiologia descritiva, confere-se maior relevância aos estudos analíticos, no sentido do seu potencial para o esclarecimento da etiologia de surtos.[5,6] Na maioria dos casos em que se denominam estudos específicos, mencionam-se os estudos de prevalência, sendo denominados *surveys* ou *encuestas* em populações ou, então, estudos de relato de caso e séries de casos clínicos. Com frequência, há inconsistência nas denominações.

Na maioria dos livros de texto brasileiros de epidemiologia, mostrados na Figura 3, incluem-se capítulos

Autores	Ano	Referência
MacMahon B, Pugh TE.	1970	Epidemiology, principles and methods. 1st ed. Boston: Little, Brown and Company.
Mausner JS, Bahn AK.	1974	Epidemiology. An Introductory text. 1st ed. Philadelphia: W. B. Saunders Company.
Goodman RA, Peavy JV.	1996[a]	Describing epidemiologic data. In: Gregg MB. Field epidemiology. 1st ed. Oxford: Oxford University Press; p. 60-80.
Buehler JW, Dicker RC.	1996[a]	Designing studies in the field. In: Gregg MB. Field epidemiology. 1st ed. Oxford: Oxford University Press; p. 81-91.
Beaglehole R, Bonita R, Kjellström T.	2001	Epidemiologia básica. 2ª ed. São Paulo: Organização Mundial da Saúde e Livraria Santos Editora.
Friis RH, Sellers TA.	2004	Epidemiology for public health practice. 3rd ed. Mississauga, ON: Jones and Bartlett Publishers.
Hulley SB, Cummings SR, Browner WS, Grady D, & Newman TB.	2008	Delineando pesquisas clínicas: uma abordagem epidemiológica. 3ª ed. Porto Alegre: Artmed.
Rothman KJ, Greenland S, Lash TL.	2008	Modern epidemiology. 3rd ed. Baltimore: Wolters Kluwer/Lippincott Williams & Wilkins; Chapter 6, Types of epidemiologic studies; p. 87-99.

a) capítulos do mesmo livro de texto.

Figura 1 – Livros de texto de epidemiologia produzidos no exterior que priorizam os estudos epidemiológicos analíticos e que não detalham estudos descritivos

de análise descritiva, sempre se dando ênfase às medidas de incidência e prevalência de agravos em uma população, de acordo com as variações segundo pessoa, lugar e tempo. Enfatiza-se também a forma de análise e apresentação de dados. Entretanto, sua nomenclatura, à semelhança dos livros publicados no exterior, não inclui especificamente os estudos descritivos. Os seis autores mostrados na Figura 3 dão ênfase aos estudos analíticos. Por exemplo, em um mesmo livro de texto, o capítulo que aborda os tipos de estudo não menciona os descritivos,[7] enquanto um capítulo específico de estudos ecológicos separa as estratégias "exploratórias" das "analíticas". Tais estratégias exploratórias poderiam ser denominadas *descritivas*.[8]

Na Figura 4, mostram-se dois livros de texto brasileiros em que há menção mais detalhada de alguns estudos descritivos. Pereira (1995, p. 278)[9] caracteriza os estudos descritivos como aqueles que não apresentam grupo controle, diferenciando-os dos estudos analíticos. O autor lista nove tipos de estudo epidemiológico, sendo os quatro primeiros descritivos: estudo de caso, série de casos, estudo de incidência e estudo de prevalência (transversal ou seccional descritivo).

Ainda no âmbito nacional, Zanetta (2004)[10] reconhece a existência de estudos descritivos em contraste com os analíticos. Para a autora, os estudos descritivos "relatam prevalências ou descrevem uma situação que parece anormal", e eles são úteis para planejamento nos serviços de saúde e para geração de hipóteses "no campo experimental". Ao listar os "delineamentos de estudos em medicina", ela inclui relato de caso, série de casos e estudo transversal ou de prevalência, citados como estudos descritivos. Os restantes são estudos analíticos.

Classificar estudos descritivos: faz sentido?

Percebe-se, pela leitura de capítulos de livros de texto, que há uma diferença na abordagem, que varia da completa omissão dos estudos descritivos à sua inclusão. Quando os estudos descritivos são explicitados, há modos diferentes de conceituá-los e de classificá-los. Uma primeira dificuldade é a tendência a nomear como descritivos apenas aquelas pesquisas que são realizadas com base em dados secundários macrodemográficos ou de populações abertas não clínicas.

Também se torna relevante diferenciar entre a técnica de coleta de dados e o tipo de estudo. Nesse sentido, é importante, neste momento, lembrar a proposta de Laurenti e colaboradores,[11] que consideram que "levantamento de dados estatísticos" é o conjunto de operações que permitem a coleta de dados que possibilitam a caracterização de um evento. Essa coleta pode ser realizada com aproveitamento de dados existentes e registrados (que o autor denomina levantamento propriamente dito); ou com dados existentes, porém não registrados, coletados por meio de entrevistas ou

Inclusão de estudos descritivos	Autores	Anos	Referências	Estudos citados	Notas
Sem capítulo separado de estudos descritivos; citam-se estudos descritivos específicos em determinados capítulos ou citam-se os estudos descritivos com outros nomes.	Gordis L. 2014.	2014	Epidemiology. 5th ed. Philadelphia, PA: Elsevier/Saunders.	O autor aborda pesquisas em âmbito clínico sem grupo de comparação que envolvem o acompanhamento de pacientes (coorte de pacientes), estudos de caso e "séries de casos".	Dois capítulos, "História natural da doença e prognóstico" e "Avaliação de medidas preventivas e terapêuticas", abordam em detalhe, respectivamente, os estudos em âmbito clínico e seus desfechos (cura, controle e óbito), o primeiro, e seus métodos de aferição (tábuas de vida e análise de sobrevida), o segundo.
	Kleinbaum DG, Kupper LL, Morgenstern H.	1982	Epidemiologic research: Principles and quantitative methods. 1st ed. New York: Van Nostrand Reinhold; Chapter 3, Types of epidemiologic research; p. 40-50.	A menção de estudos descritivos encontra-se incluída nos "estudos observacionais". Nesse contexto, é definido o estudo descritivo como aquele que é realizado "quando pouco se conhece da ocorrência, história natural ou determinantes da doença". Seus objetivos então seriam determinar a frequência ou tendência temporal da doença em uma população específica e formular hipóteses etiológicas.	Na conceitualização dos estudos epidemiológicos, os autores observam que a sua caracterização depende do nível em que se implementariam, sendo que este "nível" estaria relacionado com o período da história natural da doença, adotando-se o modelo de Leavell e Clark de níveis de prevenção. Os autores lembram que os objetivos da pesquisa epidemiológica são os seguintes: descrever, explicar, predizer e controlar.
	Abramson JH, Abramson ZH.	2008	Research methods in community medicine. Surveys, Epidemiological research, Programme evaluation, Clinical trials. 6th ed. West Sussex: John Wiley & Sons; Chapter 2, Types of investigation; p. 13-34.	Mencionam-se dois tipos de estudo sob o item *Descritptive surveys*: o longitudinal (no inglês *longitudinal*), que estuda as mudanças, tais como "crescimento e desenvolvimento de crianças, mudança na taxa de suicídio, história natural da doença ou estudo de ocorrência de novos casos de doença ou óbitos na população", e de corte seccional (*cross-sectional*). Os autores incluem uma categoria de "estudos clínicos" para se referir ao estudo de "características ou evolução de uma série de pacientes".	Para os autores, os estudos descritivos descrevem a "situação da doença" na população e sua distribuição com relação a sexo, idade, religião, entre outras variáveis. Menciona-se o termo *survey* (equivalente provavelmente a inquérito ou a sondagem); entre outras acepções, os autores atribuem o de "estudo descritivo de características populacionais", "inquérito domiciliar" ou "inquérito de campo".
	Jekel JF, Katz DL, Elmore JG.	2005	Epidemiologia, Bioestatística e medicina Preventiva. 2ª ed. Porto Alegre: Artmed; Chapter 5, Delineamentos comuns de pesquisa usados em epidemiologia; p.88-99.	Inquéritos transversais, que coletam dados da frequência de fatores de risco e prevalência de doença; inquérito por entrevista; programas de triagem em massa (rastreamento, *screening*).	Os autores apontam para a existência de "delineamentos observacionais para geração de hipóteses" sem utilizar o termo "descritivo".
	Lilienfeld DE, Stolley PD.	1994	Foundations of epidemiology. 3rd ed. Oxford: Oxford University Press.	Os autores diferenciam os estudos "demográficos" (de mortalidade ou morbidade) dos estudos "epidemiológicos". Os primeiros detectariam tendências no tempo e padrões diferentes de distribuição, de acordo com a pessoa e lugar da ocorrência de óbitos ou casos de doença.	Os estudos denominados "epidemiológicos" pelos autores correspondem, *grosso modo*, aos analíticos (sejam observacionais ou experimentais).

Figura 2 – Livros de texto de epidemiologia produzidos no exterior que incluem estudos epidemiológicos descritivos

Continua

Continuação

Inclusão de estudos descritivos	Autores	Anos	Referências	Estudos citados	Notas
Sem capítulo separado de estudos descritivos; citam-se estudos descritivos específicos em determinados capítulos ou citam-se os estudos descritivos com outros nomes.	Szklo M, Nieto FJ.	2007	Epidemiology. Beyond the basics. 2nd ed. Boston: Jones and Bartlett Publishers; Chapter 1, Basic study designs in analytical epidemiology; p. 3-43.	Os autores concebem a epidemiologia descritiva e analítica. A epidemiologia descritiva faz uso de dados disponíveis para examinar de que modo as taxas (por exemplo, de mortalidade) se comportam segundo variáveis demográficas (aquelas obtidas no censo).	Com base na detecção de distribuições não uniformes, segundo os autores, os epidemiologistas definem "grupos de alto risco" para propósitos de prevenção e também para gerar hipóteses causais.
	Fletcher RH, Fletcher SW, Wagner EH.	1996	Epidemiologia clínica: elementos essenciais. 3ª ed. Porto Alegre: Artes Médicas.	Não há seção de estudos descritivos. Porém, no seu capítulo "Estudando casos", os autores mencionam o *relato de casos*, a *série de casos* e os *estudos de caso-controle*.	Estes autores sugeriram o número de 10 pacientes como o critério que separa o relato de casos da série de casos. Ao descrever a série de casos, eles dizem: "é um estudo de um grupo maior de pacientes (por exemplo 10 ou mais) com uma doença particular".
	Olsen J, Basso O.	2015	Study Design. In: Olsen J, Greene N, Saracci R, Trichopoulos D. Teaching epidemiology. A guide for teachers in epidemiology, public health and clinical medicine. 4th ed. Oxford: Oxford University Press; 37-55.	Não há seção de estudos descritivos. Porém, no seu capítulo "*Study design*", os autores mencionam o *survey*, definindo-o como estudo de prevalência em população definida.	Mencionam-se também os estudos de tipo *case-only studies* e *case-cross over studies* sem aprofundar na sua caracterização como descritivos.
Há seção de estudos descritivos.	Hennekens CH, Buring JE.	1987	Descriptive studies. In: Hennekens CH, Buring JE, Mayrent SL. Epidemiology in medicine. 1st ed. Boston: Little, Brown and Company; p. 101-131.	Estudos de correlação (equivalentes aos ecológicos analíticos); relatos de caso; série de casos; e os *cross-sectional surveys*, que podem apenas descrever exposições e/ou desfechos, ou avançar à categoria de estudos analíticos para testar, mesmo com limitações, hipóteses de associação.	Há um capítulo dedicado aos estudos descritivos definido como os que descrevem padrões da ocorrência da doença com relação a variáveis de pessoa, lugar e tempo.
	Williams CFM, Nelson KE.	2007	Chapter 3. Study designs. In: Nelson KE, Williams CFM. Infectious epidemiology: Theory and practice. 2nd ed. Toronto: Jones and Barlett Publishers; p. 63-117.	Relatos de caso; série de casos; e estudos ecológicos (analíticos).	Os autores diferenciam duas dimensões, uma clínica e uma populacional.
	Koepsell TD, Weiss NS.	2003	Epidemiologic methods. Studying the occurrence of illness. 1st ed. New York: Oxford University Press; Chapter 5, Overview of study designs; p. 93-115.	Relatos de caso; série de casos; "estudos descritivos com base em taxas" (estudos em que se combinam os casos de uma população com denominadores dessa própria população), mediante dados coletados a partir de registros contínuos, da vigilância e seus sistemas, ou de inquéritos periódicos de saúde.	Os autores admitem a existência de estudos descritivos, cuja característica principal, segundo eles, é o fato de serem conduzidos sem uma hipótese específica. Os autores diferenciam duas dimensões, uma clínica e uma populacional.
	Colimon K-M.	1990	Fundamentos de epidemiología. 1ª ed. Madrid: Ediciones Díaz de Santos; Capítulo 6, Estudios descriptivos; p. 87-112.	Inquéritos de morbidade (*encuestas de morbilidad*); inquéritos de prevalência (*encuestas de prevalencia*); estudo de população por amostragem; estudo de um segmento ou categoria não representativa de uma população; e estudos de instituição.	Não são mencionados estudos em âmbito clínico.

Figura 2 – Livros de texto de epidemiologia produzidos no exterior que incluem estudos epidemiológicos descritivos

Autores	Ano	Referência
Forattini O.	1970	Epidemiologia geral. São Paulo: Edgard Blucher, Editora da Universidade de São Paulo.
Benseñor IM, Lotufo PA.	2005	Epidemiologia. Abordagem prática. 1ª ed. São Paulo: Sarvier Editora de Livros Médicos. Capítulo 5, Principais desenhos de estudo – conceitos gerais; p. 63-89.
Bloch KV, Coutinho ESF.	2009[a]	Fundamentos da pesquisa epidemiológica. In: Medronho RA, Bloch KV, Luiz RR, Werneck GL. Epidemiologia. 2ª ed. São Paulo: Atheneu. p.173-179.
Medronho RA.	2009[a]	Estudos ecológicos. In: Medronho RA, Bloch KV, Luiz RR, Werneck GL. Epidemiologia. 2ª ed. São Paulo: Atheneu. p.265-274.
Almeida-Filho N, Barreto ML.	2011[a]	Desenhos de pesquisa em epidemiologia. In: Almeida-Filho N, Barreto ML. Epidemiologia & Saúde. Fundamentos, métodos, aplicações. 1ª ed. Rio de Janeiro: Guanabara Koogan. Capítulo 14, Desenhos de pesquisa em epidemiologia; p. 165-174.
Santana VS, Dourado I, Ximenes R, Barreto S.	2011[a]	Modelos básicos de análise epidemiológica. In: Almeida-Filho N, Barreto ML. Epidemiologia & Saúde. Fundamentos, Métodos, Aplicações. 1ª ed. Rio de Janeiro: Guanabara Koogan. Capítulo 21, Modelos básicos de análise epidemiológica: p. 232-251.

a) capítulos do mesmo livro de texto.

Figura 3 – Livros de texto de epidemiologia produzidos no Brasil que priorizam os estudos epidemiológicos analíticos e que não detalham estudos descritivos

Autores	Anos	Referências	Estudos mencionados	Notas
Pereira MG.	1995	Epidemiologia: Teoria e Prática. 1ª ed. Rio de Janeiro: Guanabara-Koogan. Capítulo 12, p. 269-288.	O autor lista nove tipos de estudo epidemiológico, sendo os quatro primeiros descritivos: estudo de caso, série de casos, estudo de incidência e estudo de prevalência (transversal ou seccional descritivo).	Para o autor, os dois primeiros tipos não são, propriamente, estudos epidemiológicos, porque só haveria "casos" na população de estudo. Por seu turno, os dois seguintes relacionariam "casos" à respectiva população.
Zanetta DMT.	2004	Delineamento de estudos em medicina. In: Massad E, Menezes RX, Silveira PSP, Ortega NRS. Métodos Quantitativos em Medicina. 1ª ed. Barueri, SP: Editora Manole. p. 389-421.	Ao listar os "delineamentos de estudos em medicina", a autora inclui relato de caso, série de casos e estudo transversal ou de prevalência, citados como estudos descritivos. Os restantes são estudos analíticos.	Zanetta reconhece a existência de estudos descritivos. Para a autora, os estudos descritivos "relatam prevalências ou descrevem uma situação que parece anormal". Eles são úteis para o planejamento nos serviços de saúde e para a geração de hipóteses "no campo experimental".

Figura 4 – Citação de estudos epidemiológicos descritivos em livros de texto de epidemiologia produzidos no Brasil em que há algum detalhamento ou caracterização de estudos epidemiológicos descritivos

exames laboratoriais (que o autor denomina inquérito); ou com dados não existentes e gerados por uma intervenção, como os eventos adversos de uma vacina (que o autor denomina dado experimental). Ressalvamos que tipos de levantamento ou de coleta de dados não determinam o tipo de estudo epidemiológico.

Diante dessa multiplicidade de pontos de vista, é relevante o âmbito em que se realiza a pesquisa para a classificação dos estudos. Inicialmente, há um âmbito populacional ou comunitário, e surge a necessidade de compreender que qualquer classificação deve contemplar duas situações nesse domínio: (i) há estudos realizados com base em dados macroestatísticos referentes a grandes conglomerados de população, comumente secundários, elaborados com numeradores individualizados mediante a notificação do caso ou óbito e denominadores baseados em estimativas de população; e (ii) estudos realizados com dados primários em vários âmbitos comunitários (locais de trabalho, de estudo, de lazer, domicílios, creches, bibliotecas públicas, sindicatos e organizações civis da comunidade, instituições religiosas, clubes, entre

outros), em que tanto numeradores como denominadores foram abordados ou reconhecidos, e computados como indivíduos.

Independentemente desse espaço populacional ou comunitário, as instituições de atenção à saúde devem ser incluídas como um âmbito específico. A princípio, é nesse âmbito que se concentram os "casos" e se registram atendimentos, intervenções, procedimentos e desfechos. Leva-se em conta que muitas dessas instituições não fazem atividades apenas com pacientes ou pessoas que já têm uma doença ou agravo. Por exemplo, vários tipos de consulta (pré-natal, controle de crescimento e desenvolvimento, puericultura, alimentação e nutrição, fluoretação tópica dental, *check-ups* de adultos), a princípio, trabalham com pessoas "da comunidade", canalizadas ou levadas a comparecer a um lócus assistencial, sem, no entanto, serem classificadas *a priori* como "casos clínicos". Outros ambientes semelhantes podem ser constituídos por bancos de sangue ou de doação de outros fluidos, hemoderivados e órgãos. Também é relevante salientar que pesquisas em que a população de estudo é constituída por pessoas avaliadas na atenção primária se encontram muito próximas do ambiente domiciliar. Assim, um caso específico pode ser o de estudos realizados nos Serviços de Internação Domiciliar que, nesse sentido, têm características clínicas, análogas à internação hospitalar. É necessário apontar que esses estudos de âmbito clínico são realizados com dados primários finitos individualizados, operacionalizados como numeradores e denominadores claramente definidos.

Atentamos para o fato de que, tanto nos estudos de âmbito clínico como nos de âmbito populacional/comunitário, a validade externa, entendida como a capacidade de generalizar os dados a partir do conjunto de sujeitos que efetivamente compõe um determinado estudo, não constitui um determinante da classificação.

Proposta de classificação de estudos descritivos

Considerando-se a necessidade de uma classificação satisfatória, resolveu-se propor uma taxonomia que esclareça potencialidades, limitações e a consistência entre os objetivos propostos e os tipos de estudos realizados.

Existem várias possibilidades de classificação. Se o critério for a observação de uma realidade ou a aferição dos efeitos de uma intervenção conduzida pelas pessoas encarregadas da pesquisa, os estudos podem ser observacionais ou de intervenção. Se o critério for a presença ou ausência de acompanhamento de pessoas ou coletivos, os estudos seriam longitudinais (chamados também de *follow-up*) ou seccionais (em um momento no tempo). Com relação à unidade de análise corresponder a pessoas ou conglomerados, os estudos poderiam ser de base individual ou ecológicos. De acordo com o âmbito em que se estabelece a pesquisa, os estudos podem ser populacionais/comunitários ou clínicos. Esta classificação não é mutuamente excludente, isto é, um estudo pode ser de unidade de observação individual e observacional, ou individual e de intervenção, ou longitudinal e observacional, por exemplo. A proposta de classificação é mostrada na Figura 5.

1. Estudos de âmbito clínico

Os estudos de âmbito clínico proporcionam dados para se entenderem as características da história natural da doença, seu diagnóstico e desfechos após o tratamento. O âmbito clínico pode ser caracterizado como as instituições de atenção primária, secundária e terciária. Contudo, ocasionalmente, a atenção primária inclui pessoas não necessariamente com doença: gestantes em consultas pré-natais, crianças sob controle de crescimento e desenvolvimento, pessoas com problemas posturais e ortopédicos, além de usuários realizando consultas de *check-up*. Em geral, correspondem à chamada epidemiologia ao pé do leito (*bedside epidemiology*).

1.1 Relato de caso

No âmbito clínico, um primeiro propósito dos pesquisadores pode ser relatar a existência de um caso ou de um número pequeno deles. A pergunta de pesquisa poderia ser a existência ou não de um determinado agravo ou doença em um país ou comunidade, as características de um número limitado de casos clínicos, ou a maneira como transcorreu a história clínica do caso, a suspeição, impressão diagnóstica e confirmação. É de fato assim que se constituiu, durante séculos, o conhecimento acumulado de entidades nosológicas, suas características e suas respostas a tratamentos.

Âmbito do estudo		Tipo de estudo	Medidas epidemiológicas
1	Clínico		
	1.1	Relato de caso	–
	1.2	Série de casos	Proporção de casos
	1.3	Coorte descritiva clínica	Incidência de eventos, letalidade
2	Populacional/comunitário		
	2.1	Estudos descritivos observacionais de prevalência	Prevalência
	2.2	Estudos descritivos observacionais de incidência ou coorte descritiva	Incidência, mortalidade
	2.3	Estudos ecológicos descritivos	Coeficientes de incidência ou de mortalidade na população

Figura 5 – Proposta de classificação dos estudos epidemiológicos descritivos

Nesse caso, a pesquisa pode se limitar a descrever em detalhe as manifestações da doença, dados da queixa principal e da anamnese, sintomas referidos, sinais clínicos detectados, resultados de exames de laboratório e imagem e a conclusão diagnóstica.

Do ponto de vista epidemiológico, o uso pode ser muito restrito para aferição de frequência na população ou mesmo da caracterização de frequência de manifestações ou achados. No entanto, a utilidade deste tipo de estudo no âmbito clínico é alertar aos profissionais de saúde sobre a existência do evento em seu meio, para efeitos de diagnóstico e diferenciação. Tais relatos servem depois para documentar a distribuição relativa dos eventos que poderiam ser cosmopolitas ou existirem, mesmo que, em algumas regiões, sejam de rara ocorrência. Um evento raro, no entanto, pode ser o primeiro sinal de uma epidemia ou de doença emergente. Há precedentes de descrições de padrões de evolução ou distribuição divergentes que levaram à conclusão da existência de eventos novos ou comportamentos inéditos de doenças existentes. Nesse sentido, estes estudos têm relevância para a vigilância epidemiológica, porque podem revelar achados preliminares de doenças emergentes ou reemergentes que se estão espalhando em novos cenários epidemiológicos.

Exemplos desse tipo de estudo epidemiológico são os relatos de casos de sarcoma de Kaposi e *Pneumocystis jirovecii* (previamente conhecido como *P. Carinii*), que serviram como evento sentinela para a existência de imunossupressão associada à aids.[14] Outro exemplo pode ser a maneira como se comportam clinicamente novos eventos até a presente década desconhecidos, tais como Chikungunyia e Zika nos países da América Latina.[15,16] É importante assinalar que é mais indicado o termo "relato de caso" que "estudo de caso", porque essa denominação pode corresponder a outras situações de pesquisa (tanto em indivíduos como em coletivos de diversos tamanhos), no âmbito da enfermagem, psicologia, serviço social ou sociologia.

Quantos indivíduos podem ser incluídos em um relato de caso? A resposta não pode ser precisa. Fletcher et al.[17] consideravam que seria de, no máximo, 10 casos. A partir desse número, seria considerado uma série de casos ou uma coorte clínica. Realmente não haveria base estatística para dizer que, a partir de 10 casos (por exemplo, um relato de casos que inclui 11 indivíduos), a aleatoriedade diminui o suficiente para se fazer uma ideia da verdadeira frequência de um dado sintoma, sinal clínico, achado laboratorial ou de imagem.

1.2 Série de casos

Ainda no âmbito clínico, a pergunta norteadora (*research question*) pode ser conhecer o comportamento de uma entidade nosológica ou doença, sua história natural e manifestações, a distribuição dos pacientes segundo sexo, idade, raça/cor da pele, segundo o subtipo de agente etiológico, épocas e locais de maior frequência, entre outras possibilidades. Esse estudo descreve o "perfil" dos casos e pode ser chamado de "série de casos". Ele avança em relação ao simples "relato de casos", cuja importância já foi destacada. Nesta situação, porém, uma quantidade maior de observações é necessária, podendo informar qual é a proporção de indivíduos que apresentam um determinado sintoma, sinal, ou característica de laboratório ou imagem. Este tipo de estudo não tem referência

populacional e habitualmente é realizado em um serviço de atenção à saúde, frequentemente hospitalar.

Contudo, uma dimensão macro do estudo de série de casos é representada pelo conjunto de notificações de "casos" constantes dos sistemas de informação para doenças ou eventos de notificação compulsória (por exemplo, o Sistema de Informação de Agravos de Notificação – Sinan). Neste caso, fazendo as devidas considerações de sensibilidade e subnotificação, estar-se-ia utilizando o "censo" dos casos detectados, o qual aumentaria a validade externa. No entanto, estaria sujeito a limitações decorrentes dos vieses assistenciais, normalmente presentes em entidades de atenção à saúde, tais como problemas de acesso e de referência entre níveis de complexidade na atenção.

Do ponto de vista da análise epidemiológica, os dados obtidos de frequência da distribuição das características entre os doentes (casos) são pontuais e específicos para essa população de "casos". A medida utilizada é a proporção de casos, porque não se trata de uma prevalência em uma comunidade ou população aberta, e sim em uma situação muito específica na qual o numerador corresponde aos casos que têm determinada característica (por exemplo, os que referiram febre, os que são do sexo masculino, ou os que tinham determinado achado de laboratório) e o denominador corresponde ao total de pacientes. Isto, assumindo-se que não se estão incluindo dados da evolução clínica dos doentes ou casos. Um exemplo é a caracterização de 87 casos de doença pelo vírus Zika em Pernambuco, destacando as características clínicas e de imagem do acometimento neurológico.[18]

1.3 Coorte descritiva clínica

Uma outra situação de pergunta norteadora de pesquisa diz respeito à evolução clínica dos casos.

Nesta situação, documenta-se a presença de "eventos novos" como metafenômenos que vão além da própria doença, tais como complicações, aparecimento de efeitos colaterais da intervenção terapêutica, cura, sequelas, ou óbitos. Podemos chamar estes estudos de coortes descritivas clínicas ou estudos descritivos de prognóstico. Percebe-se que podemos estar documentando, nestes estudos, tanto a evolução da "história natural" de uma doença, como também os efeitos de uma intervenção, isto é, em determinados estudos clínicos realizados em um grupo de indivíduos, estes podem estar sob uma terapia padronizada e o acompanhamento verifica desfechos clínicos, tais como a cura ou eventos adversos. Isto não significa que, com este estudo, se possa aferir a eficácia da intervenção, porque para isso seriam necessários grupos de comparação, configurando-se o ensaio clínico, um tipo de estudo analítico.

Do ponto de vista da análise epidemiológica, calcula-se a frequência de aparecimento desses eventos novos, comportando-se como incidência em relação ao total de pessoas. Tais eventos incidentes podem ser positivos, como a cura ou remissão. Nos casos de óbito como evento incidente, pode ser calculada a letalidade.

Nesses estudos, como em outras estratégias de acompanhamento de pessoas, os cálculos descritos no parágrafo anterior correspondem à estratégia de coorte fechada ou fixa. Uma estratégia de análise alternativa constitui a de coorte dinâmica ou aberta, em que é possível o cálculo da densidade de incidência. Neste caso computa-se, no denominador, a soma de pessoas-tempo (por exemplo, pessoas-ano), conforme a contribuição individual ao acompanhamento na coorte. O numerador continua sendo o número de casos ou de óbitos.

É importante lembrar que os eventos "novos" documentados nesse tipo de estudo fazem parte da evolução do quadro clínico, mesmo sob intervenção terapêutica. A pergunta norteadora se refere à evolução dos casos. Dados de incidência de efeitos adversos sistematizados por pesquisadores clínicos têm sido fundamentais para estabelecer a frequência esperada de tais eventos. Um exemplo é a frequência de eventos (desfechos clínicos) da COVID-19, em que os autores documentaram, além de sintomas e sinais, a incidência de complicações que levaram à internação em unidades de terapia intensiva (UTIs), ao uso de respiradores ou ao óbito, isto é, letalidade, em 1.099 pacientes na China.[19]

2. Estudos descritivos de âmbito populacional/comunitário

Este grupo de estudos corresponde a pesquisas levadas a cabo em domicílios pertencentes a bairros, municípios, regiões ou em agregados comunitários tais como escolas, igrejas, fábricas, entre outros, onde acontece parte da atividade cotidiana da população. Todas essas entidades podem ser vistas como coletivos mais ou menos restritos. Por exemplo, uma pesquisa pode ser realizada com um número importante de indivíduos, 2 mil estudantes de um só colégio ou

operários de uma fábrica. A alternativa seria ter à disposição várias unidades de conglomerados de pessoas ou mesmo sistemas inteiros, o que tornaria necessário algum tipo de procedimento de amostragem, sendo a população-fonte (ou marco amostral) todas as unidades ou conglomerados listados: escolas ou colégios de um sistema, setores censitários, templos referentes a uma única comunidade de fé, fábricas de um ramo de atividade produtiva, um setor de atividade econômica (os transportes, por exemplo). Pressupõe-se a existência de uma lista ou cadastro a partir do qual poderia ser retirada uma amostra, mediante diversas técnicas. Em suma, há uma abordagem de pessoas "sadias" (poderiam ser portadoras de uma condição não detectada ou registrada), o que os diferencia dos estudos de âmbito clínico.

2.1 Estudos descritivos observacionais de prevalência

São estudos observacionais cujo delineamento responde à pergunta de pesquisa a respeito da existência de uma dada característica no momento em que é feita a pesquisa ou a abordagem pontual dos participantes. Corresponde a estudos seccionais ou de corte seccional, também conhecidos na literatura como inquéritos ou *surveys*, que documentam eventos existentes em um determinado momento, como casos de uma doença e fatores de risco ou proteção. Estes estudos incluem os que determinam, na população, as frequências de casos, tanto os já existentes como os novos, segundo características das pessoas ou variáveis contextuais tradicionalmente atribuídas aos indivíduos (idade, sexo, etnia, *status* socioeconômico, ocupação, situação conjugal, orientação sexual, hábitos); dos locais de ocorrência (ruas, bairros, regiões administrativas, setores censitários, áreas urbanas ou rurais, municípios, estados, países); e das épocas de ocorrência (hora, dia, mês, ano).

A pergunta de pesquisa refere-se à frequência pontual de uma doença, de um fator de risco ou de uma característica específica dessa população ou segmento comunitário. Assume-se que há um recorte momentâneo no tempo (a chamada sincronia nas ciências sociais). O nome que estamos sugerindo, estudo descritivo de prevalência, descreve o que foi feito sem necessidade de outras explicações. Outra possível denominação tem sido a de inquérito, apesar da possibilidade de se confundir o tipo de estudo com uma técnica de obtenção de dados mediante a formulação de perguntas (ato de inquirir). O termo inquérito é semelhante à denominação *encuesta*, utilizada em espanhol (*encuesta epidemiológica*), ou a *enquête*, do francês. Tais termos são usados no mesmo sentido e com a mesma possibilidade de confusão, apesar de serem utilizados amplamente para propósitos específicos (inquérito epidemiológico, inquérito nutricional, inquérito alimentar, inquérito sorológico). De fato, em alguns âmbitos acadêmicos, o termo francês tem sido traduzido como enquete. No entanto, na literatura epidemiológica francesa e franco-canadense, esse termo pode ser usado também como sinônimo de "estudo", por exemplo, em *enquête analytique* e *enquête descriptive*.[20]

Em inglês, como foi mencionado acima, utiliza-se o termo *survey*, tendo originalmente a conotação de sondagem ou prospecção, e foi muito utilizado por Paul Lazarsfeld o termo *opinion survey*, para estudos que se tornaram muito populares nos Estados Unidos após a Segunda Guerra Mundial. Outras expressões como "estudo transversal descritivo" e "estudo de corte seccional descritivo" podem ser sinônimas. Caso não especifiquem sua natureza descritiva, podem causar confusão, porque na literatura são muito utilizados para designar os correspondentes estudos analíticos de prevalência.

Do ponto de vista da análise, os estudos descritivos de prevalência utilizam como medida de frequência o cálculo da prevalência. Sua validade externa depende da estratégia de amostragem.

Para propósitos de vigilância epidemiológica, alimentar e nutricional, o Estado brasileiro tem promovido, nas últimas décadas, a realização de grandes estudos de prevalência que devem ser repetidos periodicamente, tais como a Vigilância de Fatores de Risco e Prevenção para Doenças Crônicas por Inquérito Telefônico (Vigitel), a Pesquisa Nacional de Saúde do Escolar (PeNSE), a Pesquisa Nacional sobre Acesso, Utilização e Promoção do Uso Racional de Medicamentos no Brasil (PNAUM), a Pesquisa Nacional de Saúde (PNS), entre outros. Tais estudos são realizados com base em amostras complexas e abordagem dos participantes mediante comunicação telefônica, entrevistas no domicílio ou no âmbito escolar. A princípio, nos relatórios publicados, tais investigações fornecem a possibilidade de estimar prevalências gerais e sua distribuição e, nesse sentido, são descritivas, tal como

mostra a publicação periódica que relata achados do Vigitel.[21] A partir deles, podem ser realizados trabalhos analíticos para responder a outro tipo de pergunta de pesquisa que seria resolvida mediante testes de hipóteses.

2.2 Estudos descritivos observacionais de incidência ou coorte descritiva

Trata-se de pesquisas que envolvem o seguimento ou acompanhamento de um grupo populacional para investigar o aparecimento de novos desfechos (casos, recidivas, óbitos ou outros eventos), estabelecendo uma dimensão diacrônica análoga ao estudo de coorte clínica.

Quando a pergunta norteadora se refere à frequência de novos eventos na população "sadia", uma vez acompanhada, utiliza-se a denominação de estudo descritivo de coorte. Neste caso, o seguimento determinará a frequência de casos de doença, de óbitos ou outros eventos incidentes, tais como iniciação sexual, soroconversão, recidivas, entre outros. Acompanha-se uma população com características definidas.

Do ponto de vista da análise epidemiológica, na coorte descritiva, calcula-se a frequência de aparecimento desses eventos novos, como incidência em relação ao total de pessoas da comunidade efetivamente acompanhadas (incidência acumulada). Nos casos de óbito como evento incidente, calcula-se a taxa de mortalidade. Nos dois casos, os denominadores expressam a população "em risco" de acontecer o evento do numerador. Tanto na mortalidade como na morbidade, é possível aferir a densidade de incidência, com a criação do artifício pessoa-tempo no denominador.

Apresenta-se, como exemplo, uma coorte de estudantes de escolas de ensino fundamental na Tailândia, cuja experiência é acompanhada para aferir a incidência de dengue.[22]

2.3. Estudos ecológicos descritivos

Finalmente, é necessário considerar os estudos com base em dados agregados, taxas ou proporções calculadas para um grupo populacional, em que os numeradores correspondem ao número de eventos notificados ou registrados (óbitos, casos de doença de notificação, outros eventos, acidentes, violência), e os denominadores são estimativas de população intercensitária. Neste caso, os numeradores são finitos, mas os denominadores correspondem a estimativas, e qualquer comparação torna-se difícil, tanto pelas diferenças existentes na base populacional – o que faz necessário padronizar as taxas, recorrendo a outros dados (distribuição etária, por exempo) – quanto pela dificuldade de estabelecer o *status* de exposição de indivíduos, porque se avaliam agregados. A pergunta de pesquisa seria: qual é a frequência (incidência, mortalidade) do evento em determinada população? Como evoluiu ao longo dos anos? Assim, a agregação nos leva a uma abordagem ecológica, frequentemente com base em dados secundários. Também poderiam ser utilizados para examinar, de modo ecológico, os efeitos de uma vacinação no nível da população, comparando-se a taxa de incidência antes e depois de uma intervenção.[23]

Conclusão

Os estudos descritivos têm sido relegados ao ostracismo na literatura científica geral e na epidemiológica em particular. Foi demonstrado anteriormente que eles respondem a perguntas científicas válidas e relevantes. A sobrevalorização de métodos de inferência para responder a questões de pesquisas pertinentes aos métodos da epidemiologia analítica levou a uma rejeição sistemática da epidemiologia descritiva, cujo resgate foi reivindicado no passado.[2] Ao descrever os estudos, há o cuidado de colocar as perguntas científicas às quais cada tipo responderia. Tais descrições levam à formulação de hipóteses que seriam ulteriormente testadas mediante estudos analíticos, com apoio da estatística inferencial.

Ao se mostrar que vale a pena refletir sobre o papel dos estudos descritivos e sua classificação, podem-se enxergar outras potencialidades destes estudos. Algumas técnicas que apoiam a análise descritiva podem ter sido subutilizadas, em virtude da hegemonia dos estudos analíticos inferenciais. Por exemplo, a utilização de análise fatorial de correspondência e de análise de componentes principais para descrever a agregação de indivíduos conforme variáveis, as análises geográficas com base em dados georreferenciados, apoiadas pela estatística espacial, as séries temporais e análises de sobrevida. Um maior emprego de técnicas e de métodos quantitativos enriqueceria os estudos acima classificados e contribuiria para os objetivos e a melhor utilização da epidemiologia descritiva.

Para propósitos de vigilância epidemiológica, a repetição de estudos de prevalência, em amostras obtidas a partir da mesma população-fonte ou marco amostral, pode indicar tendências na existência de desfechos em saúde-doença. Pode se tratar de hábitos ou práticas relevantes à saúde, infecção (marcadores de sorologia), ou doença. Tais estudos seriados de prevalência ou estudos de painel (*pannel studies*) utilizam a mesma população-fonte – a exemplo de detentores de linhas telefônicas, moradores cadastrados em setores censitários, estudantes matriculados em um sistema escolar, entre outros. Obviamente, com o transcorrer dos anos, a população real pode mudar (os adolescentes de um sistema escolar já não estarão na escola cinco anos depois), mas a população-alvo a ser monitorada (adolescentes) continua a ser a mesma. Por exemplo, a repetição do Vigitel, da PeNSE, da PNS, ou de outras pesquisas de base populacional, produz dados relevantes, indicando tendências ou novas hipóteses a serem testadas.

Referências

1. Porta M. A dictionary of epidemiology. 5th ed. Oxford: Oxford University Press; 2008.
2. MacMahon B, Pugh TE. Epidemiology, principles and methods. Boston: Little, Brown and Company; 1970.
3. Rothman KJ, Greenland S, Lash TL. Types of epidemiologic studies. In: Modern epidemiology. 3rd ed. Baltimore: Wolters Kluwer; Lippincott Williams & Wilkins; 2008. p. 87-99.
4. Gordis L. Epidemiology. 5th ed. Philadelphia, PA: Elsevier; Saunders; 2014.
5. Goodman RA, Peavy JV. Describing epidemiologic data. In: Gregg MB. Field epidemiology. Oxford: Oxford University Press; 1996. p. 60-80.
6. Buehler JW, Dicker RC. Designing studies in the field. In: Gregg MB. Field epidemiology. Oxford: Oxford University Press; 1996. p. 81-91.
7. Bloch KV, Coutinho ESF. Fundamentos da pesquisa epidemiológica. In: Medronho RA, Bloch KV, Luiz RR, Werneck GL. Epidemiologia. 2. ed. São Paulo: Atheneu; 2009. p.173-9.
8. Medronho RA. Estudos ecológicos. In: Medronho RA, Bloch KV, Luiz RR, Werneck GL. Epidemiologia. 2. ed. São Paulo: Atheneu; 2009. p. 265-74.
9. Pereira MG. Epidemiologia: teoria e prática. Rio de Janeiro: Guanabara-Koogan; 1995. p. 269-88.
10. Zanetta DMT. Delineamento de estudos em medicina. In: Massad E, Menezes RX, Silveira PSP, Ortega NRS. Métodos quantitativos em medicina. Barueri, SP: Editora Manole; 2004. p. 389-421.
11. Laurenti R, Mello Jorge MHP, Lebrão ML, Gotlieb SLD. Estatísticas de saúde. São Paulo: EPU; EDUSP; 1985.
12. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York: Van Nostrand Reinhold; 1982.
13. Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS, editors. Methods in social epidemiology. San Francisco: Jossey-Bass, a Wiley Imprint; 2006. p. 393-428.
14. Hymes KB, Cheung T, Greene JB, Prose NS, Marcus A, Ballard H, et al. Kaposi's sarcoma in homosexual men: a report of eight cases. Lancet [Internet]. 1981 Sep [cited 2020 Oct 13];2(8247):598-600. Available from: https://doi.org/10.1016/s0140-6736(81)92740-9
15. Sá PKO, Nunes MM, Leite IR, Campelo MDGLDC, Leão CFR, Souza JR, et al. Chikungunya virus infection with severe neurologic manifestations: report of four fatal cases. Rev Soc Bras Med Trop [Internet]. 2017 Mar-Apr [cited 2020 Oct 13];50(2):265-68. Available from: https://doi.org/10.1590/0037-8682-0375-2016
16. Torres JR, Leopoldo-Códova G, Castro JS, Rodríguez L, Saravia V, Arvelaez J, et al. Chikungunya fever: atypical and lethal cases in the Western hemisphere: a Venezuelan experience. IDCases [Internet]. 2014 Dec [cited 2020 Oct 13];2(1):6-10. Available from: https://doi.org/10.1016/j.idcr.2014.12.002
17. Fletcher RH, Fletcher SW, Wagner EH. Epidemiologia clínica: elementos essenciais. 3. ed. Porto Alegre: Artes Médicas; 1996.
18. Meneses JDA, Ishigami AC, Mello LM, Albuquerque LL, Brito CAA, Cordeiro MT, et al. Lessons learned at the epicenter of Brazil's Congenital Zika epidemic: evidence from 87 confirmed cases. Clin Infect Dis [Internet]. 2017 May [cited 2020 Oct 13];64(10):1302-08. Available from: https://doi.org/10.1093/cid/cix166

19. Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, et al. Clinical Characteristics of coronavirus disease 2019 in China. N Engl J Med [Internet]. 2020 Apr [cited 2020 Oct 13];382:1708-20. Available from: https://doi.org/10.1056/NEJMoa2002032

20. Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. Departamento de Vigilância de Agravos e Doenças Não Transmissíveis e Promoção da Saúde. Vigitel Brasil 2014: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico [Internet]. Brasília: Ministério da Saúde; 2015 [citado 2020 out 13]. 152 p. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/vigitel_brasil_2014.pdf

21. Sabchareon A, Sirivichayakul C, Limkittikul K, Chanthavanich P, Suvannadabba S, Jiwariyavej V, et al. Dengue infection in children in Ratchaburi, Thailand: a cohort study. I. Epidemiology of symptomatic acute dengue infection in children, 2006–2009. PLoS Negl Trop Dis [Internet]. 2012 Jul [cited 2020 Oct 13];6(7):e1732. Available from: https://dx.doi.org/10.1371%2Fjournal.pntd.0001732

22. Freitas HSA, Merchan-Hamann E. Impacto de la vacuna conjugada en la incidencia de meningitis por Haemophilus influenzae en el Distrito Federal de Brasil: resultados de tres años de seguimiento. Rev Pan Salud Pública [Internet]. 2006 jan [citado 2020 out 13];19(1):33-7. Disponível em: https://www.scielosp.org/article/rpsp/2006.v19n1/33-37/

23. Barata RCB. O desafio das doenças emergentes e a revalorização da epidemiologia descritiva. Rev Saúde Pública [Internet]. 1997 out [citado 2020 out 13];31(5):531-7. Disponível em: http://dx.doi.org/10.1590/S0034-89101997000600015

Abstract

Descriptive epidemiological studies are of relevance, given that there are inconsistencies in the literature with regard to their nomenclature and classification. We reviewed 19 international and six national academic textbooks on epidemiology, where the main criterion was to have them available in order to undertake an in-depth review of chapters on descriptive epidemiology and study types. In 11 books, the authors prioritize analytical studies. Twelve foreign texts and two from Brazil include descriptive studies, although the majority did not specifically refer to a category with this name. We propose a classification based on the answers to research questions, including the following types of study: case report, case series, clinical cohort, prevalence study, incidence study (cohort) and descriptive ecological study. We discuss potential uses, implementation of novel data analysis methods and their relevance in health surveillance.

Keywords: *Epidemiological Studies; Epidemiology; Descriptive; Classification.*

Resumen

La categoría de estudios epidemiológicos descriptivos es relevante para los servicios de atención de salud ya que existen inconsistencias en la literatura con relación a su nomenclatura y clasificación. Se revisaron libros de texto académicos de epidemiología con ejemplares disponibles para revisión detallada de capítulos de epidemiología descriptiva y tipos de estudio: 19 extranjeros y 6 brasileños. En 11 libros, los autores no consideran ningún estudio que no sea analítico. Doce textos extranjeros y dos brasileños abarcan estudios descriptivos, aunque la mayoría no reconozca esa categoría explícitamente. Se propone una clasificación basada en las respuestas a preguntas orientadoras de la investigación incluyendo los siguientes tipos de estudios: relato de caso, serie de casos y cohorte clínica; cuatro de ámbito poblacional/comunitario: estudio de prevalencia, estudio de incidencia (cohorte), estudio descriptivo ecológico. Se discuten las potencialidades del uso, la implementación de nuevos métodos de análisis y su relevancia en la vigilancia epidemiológica.

Palabras clave: *Estudios Epidemiológicos; Estudios Descriptivos; Clasificación.*

Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento

Article *in* Epidemiologia e Serviços de Saúde · December 2003

DOI: 10.5123/S1679-49742003000400003

2 authors, including:

Sandhi MARIA Barreto
Federal University of Minas Gerais
313 PUBLICATIONS **11,542** CITATIONS

SEE PROFILE

Artigo
de revisão

Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento

Types of Epidemiologic Studies: Basic Concepts and Uses in the Area of Aging

Maria Fernanda Lima-Costa
Núcleo de Estudos em Saúde Pública e Envelhecimento/Fiocruz, Belo Horizonte-MG
Departamento de Medicina Preventiva e Social/Universidade Federal de Minas Gerais, Belo Horizonte-MG

Sandhi Maria Barreto
Núcleo de Estudos em Saúde Pública e Envelhecimento/Fiocruz, Belo Horizonte-MG
Departamento de Medicina Preventiva e Social/Universidade Federal de Minas Gerais, Belo Horizonte-MG

Resumo

Os idosos constituem o segmento que mais cresce na população brasileira. Estudos epidemiológicos sobre as condições e determinantes da saúde do idoso são fundamentais para subsidiar políticas de saúde voltadas a essa população. No presente trabalho, são expostos alguns conceitos básicos da epidemiologia, os principais delineamentos de estudos observacionais e suas aplicações na área de envelhecimento. Os estudos descritivos e analíticos (ecológico, seccional, caso-controle e coorte) são apresentados e exemplificados com trabalhos realizados no Brasil. São discutidas as principais fontes de vieses em estudos epidemiológicos sobre envelhecimento, tais como uso de respondentes próximos, exclusão de idosos institucionalizados e o efeito de viés de sobrevivência e alguns cuidados necessários ao planejamento, condução, análise e interpretação dos resultados desses estudos.

Palavras-chave: epidemiologia; envelhecimento; delineamento de estudos; vieses.

Summary

Older adults are a population group that is increasing most rapidly in Brazil. Epidemiological studies of health conditions and determinants in the elderly are essential to help develop health policies for this population. In this work we present some basic concepts in epidemiology, the main design of observational studies, and their application in the field of aging. Descriptive and analytical studies (ecological, cross-sectional, case-control and cohort) are presented using examples of research projects carried out in Brazil. The main sources of bias, such as the use of proxy respondent, exclusion of institutionalized persons and survival bias are discussed, and some considerations are presented that must be taken into account the design, conduction, analysis and interpretation of results from these studies.

Key words: *epidemiology; aging; study design; bias.*

Endereço para correspondência:
Núcleo de Estudos em Saúde Pública e Envelhecimento/Fiocruz, Av. Augusto de Lima, 1715, Belo Horizonte-MG. CEP: 30190-002.
E-mail: lima-costa@cpqrr.fiocruz.br

Introdução

Os idosos, particularmente os mais velhos, constituem o segmento que mais cresce na população brasileira. Entre 1991 e 2000, o número de habitantes com 60-69, 70-79 e 80+ anos de idade cresceu duas a quatro vezes mais (28, 42 e 62%, respectivamente) do que o resto da população brasileira (14%).[1,2] Uma das conseqüências do crescimento da população idosa é o aumento da demanda por serviços médicos e sociais. A análise das informações existentes sobre internações hospitalares no âmbito do Sistema Único de Saúde (SUS) mostra que o envelhecimento da população não pode ser encarado somente em termos do número absoluto ou relativo da população idosa, ou das repercussões desse aumento para a previdência social. As demandas dessa população por assistência médica são tão expressivas que o seu atendimento já responde por 23% dos gastos públicos com internações hospitalares do tipo I, no país.[3]

Estudos epidemiológicos têm mostrado que doenças e limitações não são conseqüências inevitáveis do envelhecimento, e que o uso de serviços preventivos, eliminação de fatores de risco e adoção de hábitos de vida saudáveis são importantes determinantes do envelhecimento saudável.[4,5] Como pode ser visto na Tabela 1, parte expressiva das causas de mortalidade entre idosos no Brasil poderia ser reduzida com a implementação de programas de prevenção e tratamento adequados. As doenças cardiovasculares constituem o principal grupo de causas de mortalidade entre idosos, em países como os Estados Unidos da América e o Brasil.[3,5] Fatores de risco modificáveis, que são responsáveis pela morte prematura atribuída a doenças cardiovasculares entre idosos, incluem tabagismo, consumo excessivo de álcool, inatividade física, obesidade, dislipidemia e controle inadequado da hipertensão e do diabetes.[5-10] A redução do risco cardiovascular tem-se mostrado custo-efetiva e deveria ser enfatizada ao longo da vida, da infância à velhice.[5]

Pneumonia e influenza são importantes causas de hospitalização e morte entre a população idosa. Todos os idosos deveriam receber, anualmente, vacinação contra a gripe e vacinação contra pneumonia – ou, pelo menos, uma vez na vida.[5] A morbidade e a mortalidade associadas a diferentes tipos de câncer aumentam com a idade. Os cânceres de mama e da próstata são os mais freqüentes entre mulheres e homens idosos, respectivamente. A prevenção secundária, por meio da detecção precoce, é a melhor forma de redução da mortalidade associada a esses cânceres.[11] O uso de cigarro está associado a várias das principais causas de morte entre os idosos brasileiros, tais como as neoplasias malignas da traquéia, brônquios e pulmões,

Tabela 1 - Principais causas de mortalidade entre homens e mulheres idosos (60+) segundo o capítulo da CID-10* e as duas causas mais freqüentes em cada capítulo (CID 3 dígitos). Brasil, 1996

Causas	Homens		Mulheres	
	Nº de óbitos	Taxa por 100.000	Nº de óbitos	Taxa por 100.000
Capítulo IX: Doenças cardiovasculares	**90.447**	**1.599,1**	**90.975**	**1349,3**
I60 a I69 - Doenças cerebrovasculares	29.306	518,1	29.410	436,2
I20 a I25 - Doenças isquêmicas do coração	28.479	503,5	24.650	365,6
Capítulo II: Neoplasias	**35.787**	**632,7**	**27.760**	**411,7**
C33 a C34 - Maligna da traquéia, brônquios e pulmões	6.346	112,2	-	-
C61 - Maligna da próstata	5.655	100,0	-	-
C50 - Maligna da mama	-	-	3.379	50,1
C16 - Maligna do estomago	-	-	2.510	37,2
Capítulo X: Doenças do aparelho respiratório	**32.058**	**854,6**	**27.029**	**400,9**
J40 a J44 - Doenças pulmonares obstrutivas crônicas	15.481	273.4	9.336	138,5
J12 a J18 - Pneumonia	9.211	162.8	9.601	142,4

* Capítulos da Classificação Internacional de Doenças (10ª revisão)

Fonte: SIM-Datasus, 1998a (adaptado de Lima-Costa e colaboradores, 2000a)

e as doenças pulmonares obstrutivas crônicas. Dietas ricas em frutas e verduras/legumes frescos, que contêm fibras, nutrientes essenciais e vitaminas, reduzem o risco de doenças cardiovasculares e alguns tipos de câncer. Do ponto de vista da Saúde Pública, a meta é a ingestão diária de cinco ou mais porções de frutas e verduras/legumes frescos.[12]

> *Uso de serviços preventivos, eliminação de fatores de risco e adoção de hábitos de vida salutares são importantes determinantes do envelhecimento saudável.*

Informações sobre as condições de saúde dos idosos e seus determinantes, assim como suas demandas e padrões de uso de serviços de saúde, são fundamentais para orientar políticas de saúde voltadas a essa população. Estudos epidemiológicos de base populacional, ou seja, aqueles que investigam idosos residentes na comunidade, fornecem esse tipo de informação, mas ainda são raros no Brasil. Pelo nosso conhecimento, estudos com base populacional da saúde do idosos foram ou estão sendo desenvolvidos somente no Rio Grande do Sul,[13] em três grandes cidades das regiões Sudeste e Nordeste (São Paulo,[14-16] Rio de Janeiro[17] e Fortaleza[18]) e em duas pequenas cidades no interior do país (Bambuí, em Minas Gerais;[19] e Veranópolis, no Rio Grande do Sul [20]). Existe, portanto, uma evidente carência de informações sobre as condições de saúde da nossa população idosa.

No presente trabalho, serão apresentados alguns conceitos básicos da epidemiologia, suas aplicações e particularidades para o estudo dessa população e será feita uma introdução aos principais delineamentos de estudos epidemiológicos, utilizando-se exemplos de pesquisas realizadas no país.

Epidemiologia: definição e objetivos

A Epidemiologia é definida como o estudo da distribuição e dos determinantes das doenças ou condições relacionadas à saúde em populações especificadas. Mais recentemente, foi incorporada à definição de Epidemiologia a "aplicação desses estudos para controlar problemas de saúde".[21]

Estudo inclui vigilância, observação, pesquisa analítica e experimento. **Distribuição** refere-se à análise por tempo, local e características dos indivíduos. **Determinantes** são todos os fatores físicos, biológicos, sociais, culturais e comportamentais que influenciam a saúde. **Condições relacionadas à saúde** incluem doenças, causas de mortalidade, hábitos de vida (como tabagismo, dieta, atividades físicas, etc.), provisão e uso de serviços de saúde e de medicamentos. **Populações especificadas** são aquelas com características identificadas, como, por exemplo, determinada faixa etária em uma dada população.[21]

Normalmente, os estudos epidemiológicos na área do envelhecimento centram-se nos seguintes temas: investigação dos determinantes da longevidade e das transições demográfica e epidemiológica; avaliação de serviços de saúde; e investigações da etiologia e história natural das doenças/condições relacionadas à saúde comuns entre idosos.[22]

Tipos de estudos epidemiológicos

Os estudos epidemiológicos podem ser classificados em observacionais e experimentais. Os estudos experimentais fogem ao escopo deste trabalho e não serão comentados. De uma maneira geral, os estudos epidemiológicos observacionais podem ser classificados em descritivos e analíticos.

Estudos descritivos

Os estudos descritivos têm por objetivo determinar a distribuição de doenças ou condições relacionadas à saúde, segundo o **tempo**, o **lugar** e/ou as **características dos indivíduos**. Ou seja, responder à pergunta: **quando**, **onde** e **quem** adoece? A epidemiologia descritiva pode fazer uso de dados secundários (dados pré-existentes de mortalidade e hospitalizações, por exemplo) e primários (dados coletados para o desenvolvimento do estudo).

A epidemiologia descritiva examina como a incidência (casos novos) ou a prevalência (casos existentes) de uma doença ou condição relacionada à saúde varia de acordo com determinadas características, como sexo, idade, escolaridade e renda, entre outras. Quando a ocorrência da doença/condição relacionada à saúde difere segundo o tempo, lugar ou pessoa, o epidemiologista é capaz não apenas de identificar grupos de alto risco para fins de prevenção (por exemplo: na

cidade de Bambuí, verificou-se que idosos com renda familiar inferior a três salários mínimos ingeriam menos frutas e legumes frescos e praticavam menos exercícios físicos do que aqueles com renda familiar mais alta[23]), mas também gerar hipóteses etiológicas para investigações futuras.[24]

No Brasil, existem importantes bancos de dados secundários com abrangência nacional – como o Sistema de Informações sobre Mortalidade (SIM-SUS), o Sistema de Informações sobre Autorizações de Internações Hospitalares (SIH-SUS)[25-28] e a Pesquisa Nacional de Amostra Domiciliar (PNAD, 1998)[29] – que podem ser usados em estudos epidemiológicos. Os resultados apresentados na Tabela 1 constituem exemplo de um estudo descritivo utilizando dados do SIM-SUS. Outro exemplo do uso de dados secundários para estudo epidemiológico descritivo pode ser visto na Tabela 2. Nessa tabela, verifica-se que a mortalidade por doença de Chagas no Brasil vem diminuindo progressivamente, em quase todas as faixas etárias (exceto na de 70+ anos) e que o pico da mortalidade situa-se na sexta década de vida. Resultados semelhantes são encontrados quando as taxas de mortalidade são analisadas segundo coortes de nascimento. As maiores taxas de mortalidade entre as coortes mais velhas são, possivelmente, reflexo do sucesso do programa de controle da doença de Chagas no país, representando a redução da transmissão da infecção pelo *Trypanosoma cruzi* entre as coortes mais jovens.[30]

Dados secundários também têm sido utilizados para monitorar a qualidade da assistência hospitalar prestada ao idoso. Na Figura 1, pode-se verificar que a alta mortalidade entre idosos internados em uma clínica do Rio de Janeiro (que levou ao seu fechamento temporário, a partir de denúncias divulgadas pela imprensa em 1996), já vinha ocorrendo desde 1993, sendo maior que a observada em hospitais de referência em vários dos meses estudados. Esse resultado mostra que a análise adequada de dados secundários de internações hospitalares poderia ter antecipado a identificação do problema pelos órgãos competentes, evitando o excesso de mortalidade só identificado em meados de 1996.[31]

Na Figura 2, são apresentados os resultados de um estudo descritivo usando dados primários. Nesse estudo, cerca de 1.700 idosos e uma amostra representativa de indivíduos mais jovens foram entrevistados para determinadas características, entre elas o hábito de fumar. Os resultados mostram que a

Tabela 2 - Taxas de mortalidade (por milhão) por doença de Chagas segundo o ano, com as coortes de nascimento assinaladas. Brasil, 1980, 1985, 1990 e 1995

Faixa etária (em anos)	Ano de nascimento	Anos			
		1980	1985	1990	1995
0-4	1986-90	0,4	0,7	0,1	0,1
5-9	1981-85	**0,3**	0,1	0,1	0,1
10-14	1976-80	1,2	**0,9**	0,4	0,3
15-19	1971-75	5,0	3,6	**1,3**	0,9
20-24	1966-70	10,9	6,8	4,0	**3,3**
25-29	1961-65	23,5	18,0	8,3	6,2
30-34	1956-60	**45,4**	32,6	22,0	13,5
35-39	1951-55	77,9	**50,4**	35,4	24,3
40-44	1946-50	111,5	82,2	**58,2**	40,1
45-49	1941-45	143,3	120,1	86,5	**63,8**
50-54	1936-40	171,3	151,4	129,9	103,5
55-59	1931-35	**228,3**	176,2	168,3	126,4
60-64	1926-30	249,4	**243,6**	192,2	169,0
65-69	1931-35	272,6	257,4	**233,4**	200,7
≥ 70	1926-30	59,0	74,3	89,8	**88,6**

Fonte: Adaptado de Lima-Costa e colaboradores, 2002.

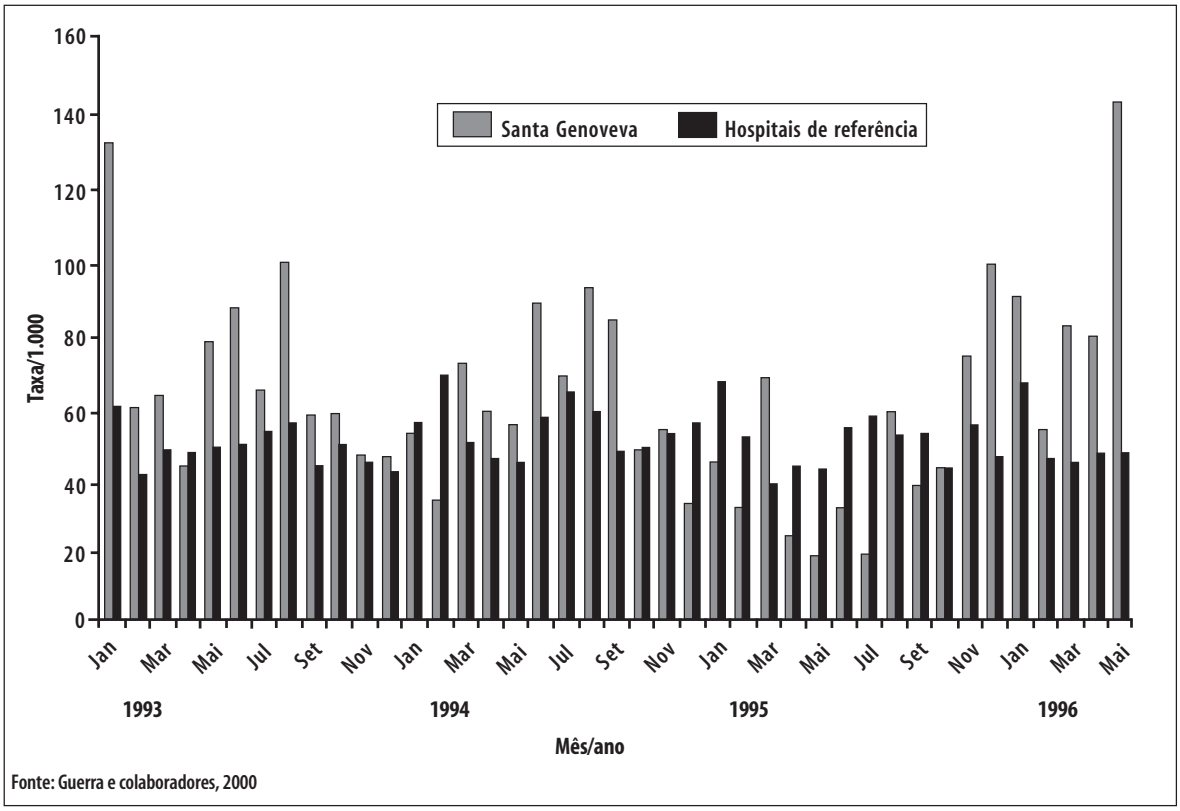


Figura 1 - Taxa de mortalidade por 1.000 entre idosos (60+) internados na Clínica Santa Genoveva, Rio de Janeiro-RJ, e entre os pacientes dos hospitais de referência. Rio de Janeiro, 1993-maio de 1996

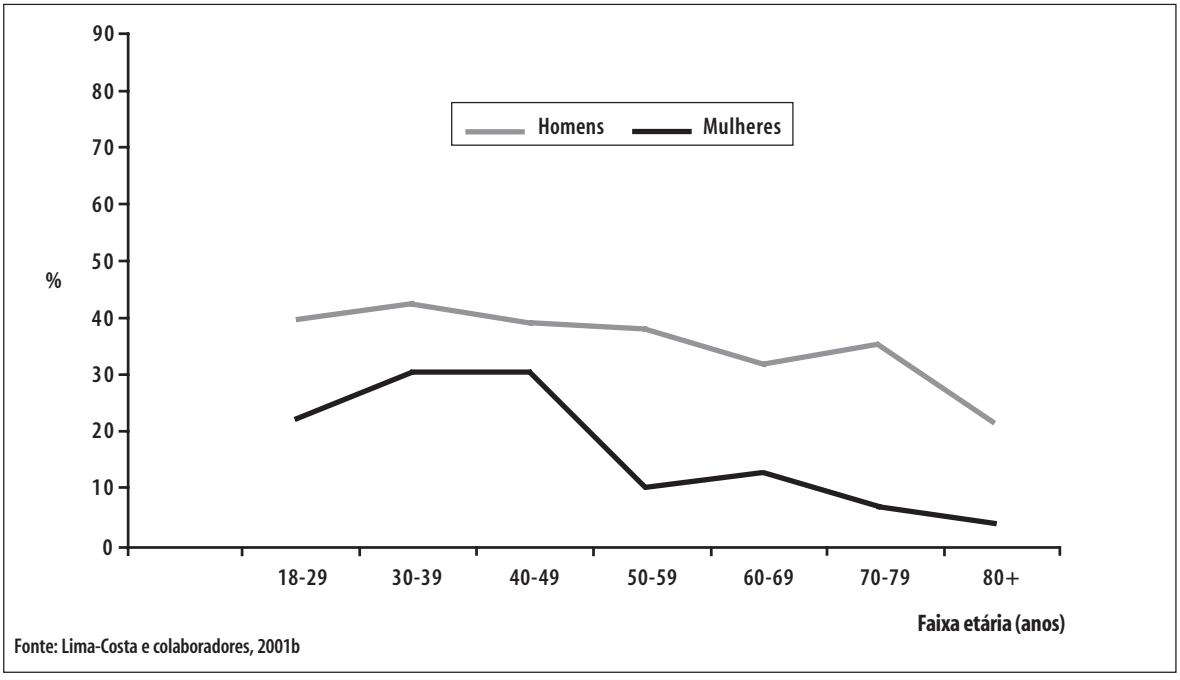


Figura 2 - Prevalência do hábito de fumar segundo o sexo e a faixa etária em Bambuí-MG. Projeto Bambuí, 1996-1997

prevalência de fumantes diminui com a idade, de forma consistente, em homens e mulheres. A redução do hábito de fumar entre pessoas mais velhas, também observada em outros trabalhos,[12] é conseqüência de pelo menos um dos seguintes fatores: a) redução do hábito de fumar em virtude do aumento da idade; b) efeito de coorte (alteração nos hábitos em gerações diferentes); e c) viés de sobrevivência (menor sobrevivência dos fumantes).[32]

Estudos analíticos

Estudos analíticos são aqueles delineados para examinar a existência de associação entre uma exposição e uma doença ou condição relacionada à saúde. Os principais delineamentos de estudos analíticos são: a) ecológico; b) seccional (transversal); c) caso-controle (caso-referência); e d) coorte (prospectivo). Nos estudos ecológicos, tanto a exposição quanto a ocorrência da doença são determinadas para grupos de indivíduos. Nos demais delineamentos, tanto a exposição quanto a ocorrência da doença ou evento de interesse são determinados para o indivíduo, permitindo inferências de associações nesse nível. As principais diferenças entre os estudos seccionais, caso-controle e de coorte residem na forma de seleção de participantes para o estudo e na capacidade de mensuração da exposição no passado, como será visto a seguir.

Estudos ecológicos

Nos estudos ecológicos, compara-se a ocorrência da doença/condição relacionada à saúde e a exposição de interesse entre agregados de indivíduos (populações de países, regiões ou municípios, por exemplo) para verificar a possível existência de associação entre elas. Em um estudo ecológico típico, medidas de agregados da exposição e da doença são comparadas. Nesse tipo de estudo, não existem informações sobre a doença e exposição do indivíduo, mas do grupo populacional como um todo. Uma das suas vantagens é a possibilidade de examinar associações entre exposição e doença/condição relacionada na coletividade. Isso é particularmente importante quando se considera que a expressão coletiva de um fenômeno pode diferir da soma das partes do mesmo fenômeno. Por outro lado, embora uma associação ecológica possa refletir, corretamente, uma associação causal entre a exposição e a doença/condição relacionada à saúde, a possibilidade do viés ecológico é sempre lembrada como uma limitação para o uso de correlações ecológicas. O viés ecológico – ou falácia ecológica – é possível porque uma associação observada entre agregados não significa, obrigatoriamente, que a mesma associação ocorra em nível de indivíduos.[24,33]

Na Figura 3, é apresentada a distribuição da proporção de óbitos por causas mal definidas entre idosos e a taxa de pobreza (proporção da população com renda *per capita* inferior a meio salário mínimo), segundo a macrorregião brasileira. Sabe-se que, para o conjunto da população idosa brasileira, cerca de 65% dos óbitos sem causa básica conhecida ocorrem sem assistência médica.[3] Assim, a maior proporção de mortes por causas mal definidas nas regiões com maior proporção de habitantes com renda familiar *per capita* inferior a meio salário mínimo sugere que a falta da assistência médica ao idoso está associada à pobreza.

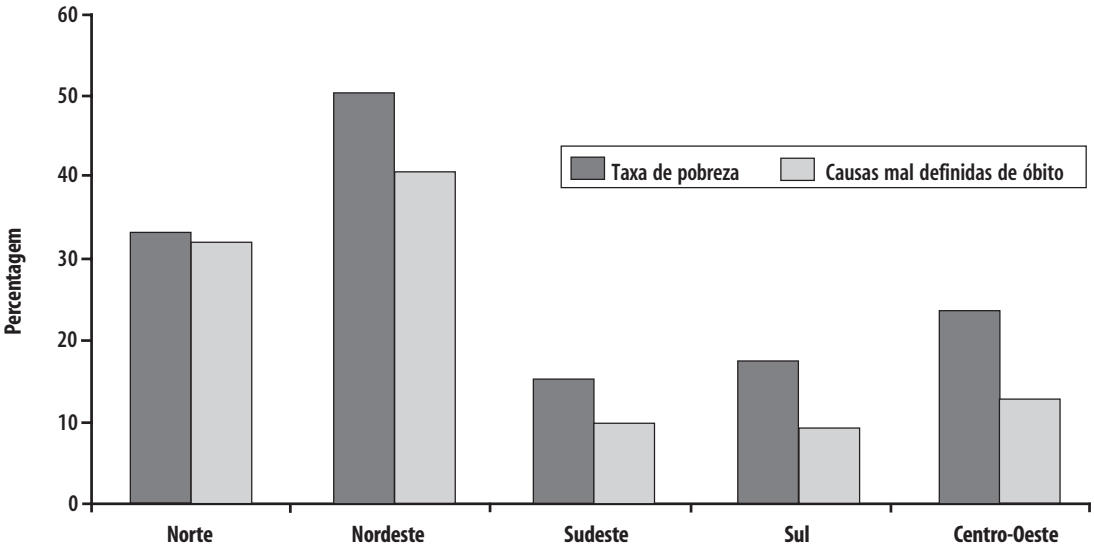
Estudos seccionais

Nos estudos seccionais, a exposição e a condição de saúde do participante são determinadas simultaneamente. Em geral, esse tipo de investigação começa com um estudo para determinar a prevalência de uma doença ou condição relacionada à saúde de uma população especificada (por exemplo, habitantes idosos de uma cidade). As características dos indivíduos classificados como doentes são comparadas às daqueles classificados como não doentes.

Um exemplo de estudo seccional foi desenvolvido na cidade de Bambuí, situada no interior de Minas Gerais, para determinar a prevalência e os fatores sociodemográficos associados à depressão.[34] Um questionário foi aplicado para identificar os indivíduos com depressão em uma amostra representativa da população da cidade com 18+ anos de idade (1.041 participantes). Os episódios depressivos atuais estavam associados ao sexo (maior prevalência no sexo feminino, em comparação ao sexo masculino), à idade (maior prevalência nos mais velhos, em comparação aos mais jovens) e à condição atual de trabalho (maior prevalência entre aqueles que não estavam trabalhando, em comparação aos que estavam), conforme

discriminação na Tabela 3. Saliente-se que as determinações do episódio depressivo atual e da ocupação foram feitas simultaneamente, ou seja, não foi possível saber se a ausência de trabalho foi anterior ou posterior ao surgimento do episódio depressivo.

Esta é a característica fundamental de um estudo seccional: não é possível saber se a exposição antecede ou é conseqüência da doença/condição relacionada à saúde. Portanto, esse delineamento é fraco para determinar associações do tipo causa-efeito, mas



Fonte: Ministério da Saúde, Brasil, 1997 a,b

Figura 3 - Proporção de óbitos por causas mal definidas entre idosos (60+) e taxa de pobreza segundo a macrorregião brasileira, 1997

Tabela 3 - Fatores sociodemográficos, independentemente associados à depressão nos últimos 30 dias determinada pelo *Composite International Diagnostic Interview* (CIDI). Projeto Bambuí, 1996-1997

Características	Depressão		OR (IC95%)
	Presente (n=85)%	Ausente (n=956)%	
Sexo			
Masculino	21,2	45,4	1,0
Feminino	78,8	54,5	2,4 (1,3-4,2)
Faixa etária (anos)			
18-29	12,9	30,0	1,0
30-44	17,7	33,7	1,2 (0,6-2,8)
45-59	36,5	22,6	3,5 (1,7-7,2)
60+	32,9	13,7	4,0 (1,9-8,5)
Situação atual de trabalho			
Trabalhando	28,2	58,7	1,0
Não trabalhando	71,8	41,3	2,1 (1,2-3,6)

Fonte: Adaptado de Vorcaro e colaboradores, 2001

* OR (IC95%): *Odds Ratio* e Intervalo de Confiança ao nível de 95%, ajustado pelas variáveis listadas na tabela, segundo o método de regressão logística. Essa é uma medida da força de associação entre variáveis (quanto maior o seu valor, maior a força da associação) (ver Tabela 4)

adequado para identificar pessoas e características passíveis de intervenção e gerar hipóteses de causas de doenças. Em relação ao estudo de Bambuí, os resultados mostraram que a depressão é um importante problema de saúde na comunidade, especialmente entre mulheres, pessoas mais velhas e aqueles que não estão trabalhando. O resultado do estudo também gerou uma hipótese sobre a influência da ausência de trabalho no desenvolvimento do episódio depressivo.

Estudos caso-controle

Os estudos caso-controle e os estudos de coorte podem ser utilizados para investigar a etiologia de doenças ou de condições relacionadas à saúde entre idosos, determinantes da longevidade; e para avaliar ações e serviços de saúde. Os estudos de coorte também podem ser utilizados para investigar a história natural das doenças.

Nos estudos caso-controle, **primeiramente**, identificam-se indivíduos com a doença (casos) e, para efeito de comparação, indivíduos sem a doença (controles) (Tabela 4). **Depois**, determina-se (mediante entrevista ou consulta a prontuários, por exemplo) qual é a Odds da exposição entre casos (a / c) e controles (b / d). Se existir associação entre a exposição e a doença, espera-se que a Odds da exposição entre casos seja maior que a observada entre controles, além da variação esperada devida ao acaso.

Tabela 4 - Delineamento de um estudo caso-controle

	Primeiramente, selecionam-se	
Depois, verifica-se a ocorrência da exposição no passado	Doentes (casos)	Não doentes (controles)
Presente	**a**	**b**
Ausente	**c**	**d**
Total	**a + c**	**b + d**

A força da associação, nesse tipo de estudo, é dada pelo Odds Ratio (OR), que é definido como a Razão de Odds – número de casos expostos sobre número de casos não expostos, dividido pelo número de controles expostos sobre o número de controles não expostos.

A fórmula para o cálculo do Odds Ratio nesta tabela é: $\frac{a}{c} / \frac{b}{d} = \frac{ad}{bc}$

Os estudos caso-controle, ao contrário dos estudos de coorte (ver a seguir), partem do efeito (doença) para a investigação da causa (exposição). Nesse artifício, residem as forças e as fraquezas desse tipo de estudo epidemiológico. Entre as vantagens, podemos mencionar: a) tempo mais curto para o desenvolvimento do estudo, uma vez que a seleção de participantes é feita após o surgimento da doença; b) custo mais baixo da pesquisa; c) maior eficiência para o estudo de doenças raras; d) ausência de riscos para os participantes; e) possibilidade de investigação simultânea de diferentes hipóteses etiológicas. Por outro lado, os estudos caso-controle estão sujeitos a dois principais tipos de vieses (erro sistemático no estudo): de seleção (casos e controles podem diferir sistematicamente, devido a um erro na seleção de participantes); e de memória (casos e controles podem diferir sistematicamente, na sua capacidade de lembrar a história da exposição). Essas limitações podem ser contornadas no delineamento e condução cuidadosos de um estudo caso-controle.[35]

Um estudo caso-controle para investigar a associação de quedas entre idosos e uso de medicamentos está sendo desenvolvido no Município do Rio de Janeiro. Os casos são pessoas com 60+ anos de idade, internadas em seis hospitais do município por fratura decorrente de queda. Os controles são pacientes dos mesmos hospitais internados por outras causas. Até o momento, os resultados sugerem um maior risco de quedas e fraturas entre aqueles que fazem uso de benzodia-zepínicos (Odds Ratio-OR=1,9; Intervalo de Confiança-IC em nível de 95%=1,0-3,8) e miorrelaxantes (OR=1,9; IC95%=1,0-4,0).[36]

Estudos de coorte

Nos estudos de coorte, **primeiramente**, identifica-se a população de estudo e os participantes são classificados em expostos e não expostos a um determinado fator de interesse (Tabela 5). **Depois**, os indivíduos dos dois grupos são acompanhados para verificar a incidência da doença/condição relacionada à saúde entre expostos (a / a + d) e não expostos (c / c + d). Se a exposição estiver associada à doença, espera-se que a incidência entre expostos seja maior do que entre não expostos, além da variação esperada devida ao acaso. Nesse tipo de estudo, a mensuração da exposição antecede o desenvolvimento da doença, não sendo sujeita ao viés

Tabela 5 - Delineamento de um estudo de coorte

Primeiramente, verifica-se a ocorrência da exposição	**Depois,** verifica-se a incidência da doença		
	Desenvolveu a doença	Não desenvolveu a doença	**Total**
Exposto	**a**	**b**	**a + b**
Não exposto	**c**	**d**	**c + d**

A força da associação, nesse tipo de estudo, é dada pelo risco relativo que é definido como a razão de incidências entre expostos e não expostos.

A fórmula para o cálculo do risco relativo nesta tabela é: $\frac{a/a+b}{c/c+d}$

de memória como nos estudos caso-controle. Além disso, os que desenvolveram a doença e os que não desenvolveram não são selecionados, mas sim identificados dentro das coortes de expostos e não expostos, não existindo o viés de seleção de casos e controles. Os estudos de coorte permitem determinar a incidência da doença entre expostos e não expostos e conhecer a sua história natural.

A principal limitação para o desenvolvimento de um estudo de coorte, além do seu custo financeiro, é a perda de participantes ao longo do seguimento por conta de recusas para continuar participando do estudo, mudanças de endereços ou emigração. Os custos e as dificuldades de execução podem comprometer o desenvolvimento de estudos de coorte, sobretudo quando é necessário um grande número de participantes ou longo tempo de seguimento para acumular um número de doentes ou de eventos que permita estabelecer associações entre exposição e doença.[37]

Por essas razões, são poucos os estudos de coorte com base populacional desenvolvidos entre idosos brasileiros. Um desses estudos (Epidoso) está sendo desenvolvido na cidade de São Paulo, onde cerca de 1.700 pessoas com 65+ anos estão sendo acompanhadas.[16] Um outro estudo (Projeto Bambuí) está sendo desenvolvido na cidade de Bambuí, Minas Gerais, onde estão sendo acompanhados todos os residentes na comunidade com 60+ anos de idade (cerca de 1.700 pessoas).[19] De uma maneira geral, os principais objetivos de um estudo prospectivo consistem em determinar a incidência de condições adversas à saúde e investigar determinantes dessas condições.

O delineamento básico do estudo de coorte de Bambuí está apresentado na Figura 4. Inicialmente, foi conduzido um censo para identificar todos os residentes na cidade. Em seguida, aqueles com 60+ anos de idade foram convidados a participar do estudo. Os que aceitaram participar foram incluídos na linha de base do estudo e submetidos a entrevista, exame físico e diversos exames laboratoriais. A entrevista foi realizada com a utilização de um questionário estruturado e pré-codificado, contendo informações sobre características sociodemográficas, morbidade auto-referida, uso de medicação, uso de serviços de saúde e fontes de cuidados, hábitos de vida, aspectos psicossociais, história reprodutiva, função física e saúde mental. Foi constituída uma soroteca e um banco de DNA para investigações futuras. As informações obtidas na linha de base do estudo são denominadas variáveis exploratórias (exposição) e a sua associação com condições adversas de saúde (variáveis de desfecho) serão investigadas, comparando-se as incidências dessas condições ao longo do tempo, entre expostos e não expostos. As principais variáveis de desfecho investigadas nesse estudo são: morte; internações hospitalares; declínio físico e cognitivo; acidentes; episódios depressivos; e uso de medicamentos e de serviços de saúde. A adesão ao estudo foi alta, tanto na linha de base (dos 1.742 idosos selecionados, 92% foram entrevistados e 86% examinados) quanto no primeiro seguimento (somente 1,7% foram perdidos para acompanhamento). Esses resultados mostram que a escolha da cidade e a forma de abordagem da comunidade foram adequadas para garantir a adesão ao estudo na linha de base e a pequena perda de acompanhamento, condição essencial para o sucesso de um estudo de coorte.[19]

Nas últimas décadas, importantes estudos de coorte com base populacional de idosos vêm sendo realizados em países desenvolvidos.[38-46] Os resultados dessas pesquisas têm sido fundamentais para subsidiar programas de prevenção e promoção da saúde dessas populações. Não se sabe, entretanto, se esses resultados são generalizáveis para países em desenvolvimento. Estudos de coorte com base populacional da população idosa nesses países são importantes para, entre outras razões: a) determinar a incidência de eventos adversos de saúde entre idosos, orientando estratégias de prevenção adequadas à realidade nacional; b) contribuir para o entendimento da etiologia de algumas doenças; e c) estudar fatores

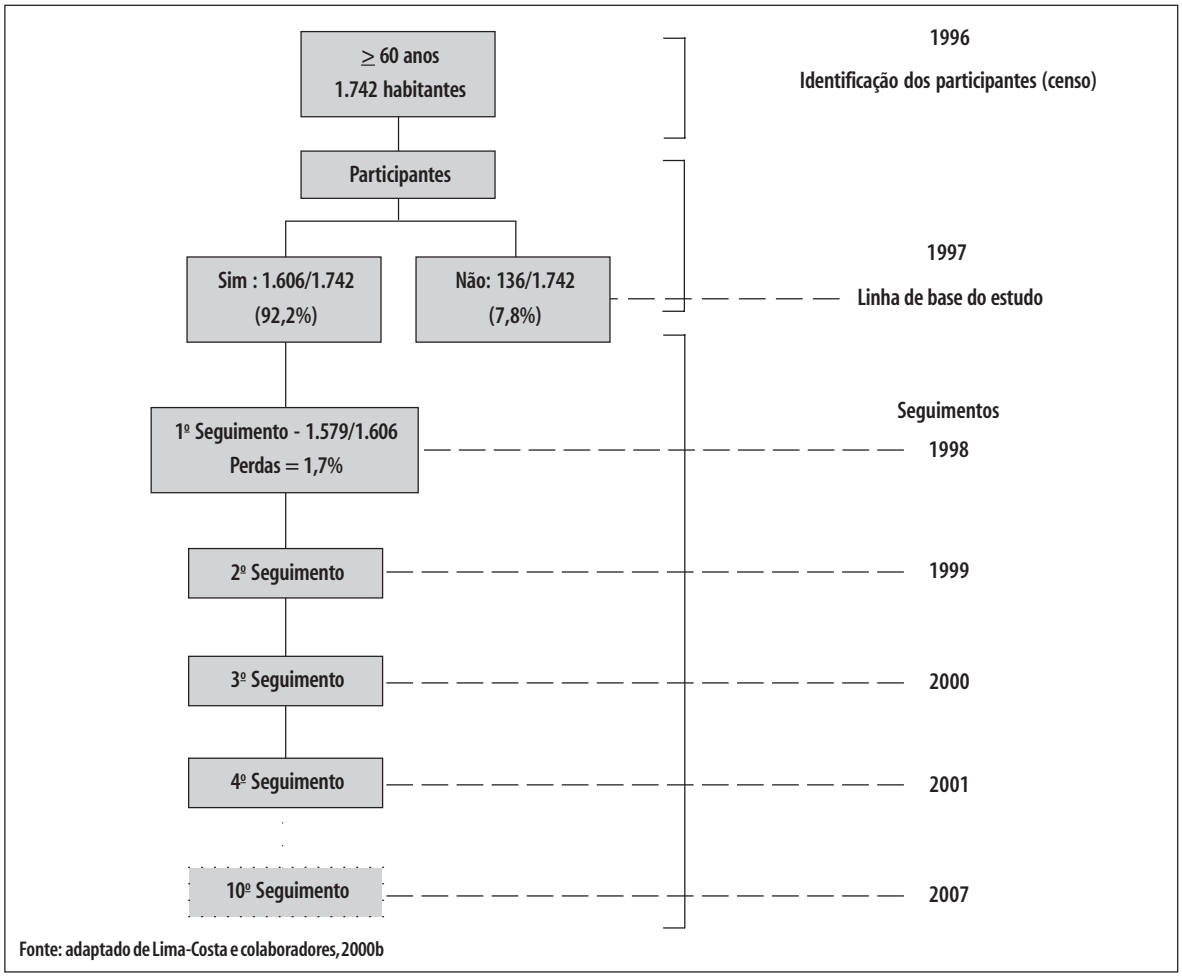


Figura 4 - Delineamento do estudo de coorte de Bambuí-MG. Projeto Bambuí, 1996-2007

culturais, comportamentos e estilos de vida que podem variar entre comunidades e países, associados a esses eventos.[19]

Vieses e variáveis de confusão

Além dos aspectos gerais da pesquisa epidemiológica, os estudos sobre envelhecimento requerem alguns cuidados ou estratégias especiais a serem levados em conta, tanto no planejamento quanto na condução, análise e interpretação dos resultados.[22,47] Entre esses aspectos, destaca-se o uso de respondentes próximos. Alguns idosos mais velhos podem estar muito doentes ou apresentar déficit cognitivo que impeça a sua participação na pesquisa. Nesse caso, pode-se recorrer a uma pessoa próxima para se obter alguma informação e assegurar a validade interna do estudo. É fundamental, entretanto, que o uso de respondente próximo seja considerado na análise (mediante estratificação ou ajustamento, por exemplo) e na interpretação dos resultados da pesquisa.[48]

Uma dificuldade dos estudos epidemiológicos sobre envelhecimento é a definição da população-alvo. Isso é particularmente importante quando o estudo inclui idosos mais velhos, porque a institucionalização cresce de forma marcante com a idade. Estudos epidemiológicos de idosos residentes na comunidade, que excluem idosos institucionalizados, podem subestimar a prevalência de incapacidade na população. Esse viés será mais acentuado em comunidades com maior grau de institucionalização.

O viés de seleção sempre deve ser lembrado em estudos do tipo caso-controle da população idosa. Ele ocorre quando casos e controles diferem entre si

sistematicamente, devido à forma de seleção. O recruta-mento de casos entre pacientes hospitalizados (ou institucionalizados) é particularmente sujeito ao viés de seleção, porque os fatores que levam à hospitalização – por exemplo: gravidade da doença, tabagismo e maior idade – também estão associados a muitos fatores de risco.[22]

O viés de sobrevivência, igualmente, deve ser considerado em estudos sobre a saúde do idoso. Os participantes idosos de estudos epidemiológicos são sobreviventes porque aqueles expostos a fatores de risco têm maior probabilidade de morte prematura. Esse viés tende a reduzir a magnitude das associações encontradas entre fatores de risco e doença/condição relacionada à saúde entre idosos.[19]

Para finalizar, também é importante considerar o efeito de variáveis de confusão nos estudos epidemiológicos sobre envelhecimento, ou seja, de fatores que podem ser uma explicação alternativa para a associação encontrada.[24,35,37] O fator de confusão está presente quando duas variáveis são associadas, mas parte da associação – ou toda ela – é decorrente de uma associação independente com uma terceira variável (de confusão). Por exemplo, as quedas podem estar associadas ao uso de diuréticos, sugerindo um efeito causal. A insuficiência cardíaca, entretanto, confunde esta associação porque o uso de diuréticos faz parte do seu tratamento e a insuficiência cardíaca é também um fator de risco para quedas.[22] O efeito de confusão pode ser controlado mediante estratificação ou ajustamento na análise dos dados.

A idade é um fator potencial de confusão de muitas associações porque, freqüentemente, está associada à exposição e à doença/condição em diferentes situações. O efeito da idade pode ser controlado mediante pareamento, estratificação ou ajustamento na análise. Quando o estudo inclui idosos mais velhos, recomenda-se o ajustamento pela idade com intervalos mais curtos (ou como variável contínua), em lugar de intervalos mais amplos (cinco em cinco ou dez em dez anos, por exemplo).[22]

Conclusões

Este trabalho apresenta, de forma sucinta, alguns conceitos básicos da epidemiologia e os delineamentos de estudos epidemiológicos observacionais que podem ser utilizados para a investigação de doenças e fatores associados a elas na população idosa. Além dos aspectos abordados, é importante lembrar que o desenvolvimento de um estudo epidemiológico envolve, pelo menos, seis etapas:

1. definição dos objetivos;
2. escolha do delineamento adequado, segundo a viabilidade do estudo e os recursos disponíveis;
3. identificação da população de estudo;
4. planejamento e condução da pesquisa;
5. coleta, análise e interpretação dos dados; e
6. divulgação dos resultados.

A qualidade de um estudo epidemiológico depende, entre outros fatores, da representatividade dos participantes, da qualidade da informação sobre a exposição e a doença/condição relacionada à saúde, da ausência de vieses e do controle adequado das variáveis de confusão. Portanto, antes de iniciar uma pesquisa, é preciso definir, cuidadosamente, a população de estudo, o tamanho da amostra (quando for o caso) e o método de seleção dos participantes. Os instrumentos de coleta de dados devem ser desenvolvidos e pré-testados, tendo em vista o conjunto de informações ou medidas que se deseja obter.

Para o desenvolvimento de um estudo epidemiológico, é preciso considerar as questões éticas pertinentes. No Brasil, aprovou-se, recentemente, um conjunto de normas éticas a serem observadas na condução de estudos envolvendo seres humanos.[49] Por exigência dessas normas, os protocolos para desenvolvimento de estudos epidemiológicos utilizando dados primários devem ser aprovados por um comitê de ética credenciado.

O envelhecimento das populações é um dos mais importantes desafios para a Saúde Pública contemporânea, especialmente nos países em desenvolvimento, onde o envelhecimento ocorre em um ambiente de pobreza e grande desigualdade social. Estudos epidemiológicos de boa qualidade e delineados de forma a contemplar essas especificidades são essenciais para subsidiar o desenvolvimento de políticas de saúde adequadas à realidade da população de idosos nesses países, para que envelheçam com saúde.

Referências bibliográficas

1. Instituto Brasileiro de Geografia e Estatística. Censo demográfico de 1991. Rio de Janeiro: IBGE; 1992.
2. Instituto Brasileiro de Geografia e Estatística. Censo demográfico de 2000. Rio de Janeiro: IBGE; 2002.
3. Lima-Costa MFF, Guerra HL, Barreto SM, Guimarães RM. Diagnóstico de saúde da população idosa brasileira: um estudo da mortalidade e das internações hospitalares públicas. Informe Epidemiológico do SUS 2000;9(1):23-41.
4. Kannel WB. Cardiovascular risk factors in the elderly. Coronary Artery Disease 1997;8/9:566-575.
5. Desai MM, Zhang P, Hennessy CH. Surveillance for morbidity and mortality among older adults – United States, 1995-1996. Morbidity and Mortality Weekly Report 1999;48(SS-8):7-25.
6. Dufour M, Fuller RK. Alchohol in the elderly. Annual Review Medicine 1995;46:123-132.
7. Corti MC, Guralnik JM, Bilato C. Coronary heart disease risk factors in older persons. Aging 1996;8:75-89.
8. Tresch DD, Aronow WS. Smoking and coronary artery disease. Clinical Geriatric Medicine 1996;12:23-32.
9. Cherubini A, Lowenthal DT, Williams LS, Maggio D, Mecocci P, Senin U. Physical inactivity and cardiovascular health in the elderly. Aging 1998;10:13-25.
10. Samos LF, Roos BA. Diabetes *mellitus* in older persons. Medical Clinics of North America 1998;82:791-803.
11. Janes GR, Blackman DK, Bolen JC, Kamimoto LA, Rhodes L, Caplan LS, Nadel MR, Tomar SL, Lando JF, Greby SM, Singleton JA, Strikas RA, Wooten KG. Surveillance for use of preventive health-care services by older adults, 1995-1997. Morbidity and Mortality Weekly Report 1999;48(SS-8):51-88.
12. Kamimoto LA, Eston NA, Maurice E, Husten CG, Macera CA. Surveillance for five health risks among older adults – United States (1993-1997). Morbidity and Mortality Weekly Report 1999;48(SS-8):89-130.
13. Conselho Estadual do Idoso. Secretaria de Estado do Trabalho, Cidadania e Assistência Social. Governo do Estado do Rio Grande do Sul. Os idosos do Rio Grande do Sul. Estudo multidimensional de suas condições de vida. Relatório Preliminar de Pesquisa. Porto Alegre; 1996.
14. Ramos LR. Growing old in São Paulo, Brazil: assessment of health status and social support of elderly people from different socioeconomic strata living in the community [PhD Thesis]. London: University of London; 1987.
15. Ramos LR, Rosa TE, Oliveira ZM, Medina MC, Santos FR. Perfil do idoso em área metropolitana na Região Sudeste do Brasil: resultado de inquérito domiciliar. Revista de Saúde Pública 1993;27:87-94.
16. Ramos LR, Toniolo J, Cendoroglo MS, et al. Two-year follow-up study of elderly residents in S. Paulo, Brazil: methodology and preliminary results. Revista de Saúde Pública 1998;32:397-407.
17. Veras R. A Survey of the health of elderly people in Rio de Janeiro, Brazil [PhD Thesis]. London: University of London; 1992.
18. Coelho Filho JM, Ramos LR. Epidemiologia do envelhecimento no nordeste do Brasil: resultados de inquérito domiciliar. Revista de Saúde Pública 1993;33:445-453.
19. Lima-Costa MFF, Uchôa E, Guerra HL, Firmo JOA, Vidigal PG, Barreto SM. The Bambuí Health and Ageing Study (BHAS). Methodological approach and preliminary results of a population-based cohort study of the elderly in Brazil. Revista de Saúde Pública 2000;34:126-135.
20. Moriguchi E, Da Cruz I, Xavier F, Schwanke C, Moriguchi Y. Cognitive performance and epolipoprotein E genotype among the very elderly living in the community. 17th World Congress of the International Association of Gerontology. Gerontology 2001;47(suppl 1):54.
21. Last JM. A Dictionary of Epidemiology. 3rd ed. Oxford: Oxford University Press; 1995.
22. Ebrahim S. Principles of epidemiology in old age. In: Ebrahim S, Kalache A. Epidemiology in old age. London: BMJ Publishing Group; 1996. p.12-21.
23. Lima-Costa MFF, Uchoa E, Barreto SM, Firmo JOA, The Bhas Group. Socioeconomic position and health in a population of Brazilian elderly: The Bambuí Health and Ageing Study (BHAS). Revista Panamericana de Salud Pública 2003 jun;13(6):387-394.
24. Szklo M, Javier Nieto F. Basic study designs in analytical epidemiology. In: Szklo M, Javier Nieto F. Epidemiology: beyond the basics. Gaithersburg: Aspen Publishers Inc; 2000. p.3-51.
25. Ministério da Saúde. Indicadores e dados básicos Brasil; 1997a. Brasília: MS; 1998. Disponível em: http://www.datasus.gov.br/cgi/idb97/socio/b05.htm.

26. Ministério da Saúde. Dados de mortalidade – CID 10; 1997b. Brasília: MS; 1998. Disponível em: http://www.datasus.gov.br/tabcgi.exe?sim/dybr.def

27. Ministério da Saúde. Secretaria de Informática. Sistema de Informações sobre Mortalidade (SIM), 1979-1996 [dados em CD-ROM]. Brasília: MS; 1998a.

28. Ministério da Saúde. Secretaria de Informática. Movimento de Autorizações de Internações Hospitalares, 1995-1997 [dados em CD-ROM]. Brasília: MS; 1998b.

29. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional de Amostra Domiciliar (PNAD). Rio de Janeiro: IBGE; 1998.

30. Lima-Costa MFF, Barreto SM, Guerra HL. Chagas'disease among older adults: branches or mainstream of the present burden of *Trypanosoma cruzi* infection. International Journal of Epidemiology 2002;31(3):688-689.

31. Guerra HL, Barreto SM, Uchoa E, Firmo JOA, Lima-Costa MFF. A morte de idosos na Clínica Santa Genoveva, Rio de Janeiro: um excesso de mortalidade que o sistema público de saúde poderia ter evitado. Cadernos de Saúde Pública 2000;16:545-551.

32. Lima-Costa MFF, Barreto SM, Uchoa E, Firmo JOA, Vidigal PG, Guerra HL. The Bambuí Health and Ageing Study (BHAS): prevalence of risk factors and preventive health-care use. Revista Panamericana de Salud Pública 2001;9:219-227.

33. Morgenstern H. Ecologic studies. In: Rothman KJ, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott-Raven; 1998. p.459-480.

34. Vorcaro CMR, Lima-Costa MFF, Barreto SM, Uchoa E. Unexpected high prevalence of 1-month depression in a small Brazilian community: the Bambui Study. Acta Psychiatrica Scandinavica 2001;104:257-263.

35. Breslow NE, Day NE. Statistical Methods in Cancer Research. I. The design and analysis of case control studies. Lyon: IARC Scientific Publications; 1980.

36. Coutinho ESF, Silva SD. O uso de medicamentos e quedas entre idosos – um estudo de caso-controle na cidade do Rio de Janeiro. Anais do XII Congresso Brasileiro de Geriatria e Gerontologia e I Encontro Brasil-França de Geriatria; 2000 jun 21-24; Brasília, Brasil. Rio de Janeiro; 2000. p.74

37. Breslow NE, Day NE. Statistical Methods in Cancer Research. II. The Design and analysis of cohort studies. Lyon: IARC Scientific Publications; 1987.

38. Gural JM, Kaplan GA. Predictors of healthy ageing: prospective evidence from the Alameda Country Study. American Journal of Public Health 1979;79:703-708.

39. Coroni-Huntley JC, Brock DB, Ostfeld AM, Taylor JO, Wallace RB, editors. Established populations for epidemiological studies of the elderly. Washington (DC): National Institute on Ageing; 1986. NIH Publications, 86-2443.

40. Schoenbach VJ, Kaplan BH, Fredman L, Kleinbaum DG. Social ties and mortality in Evans Country, Georgia. American Journal of Epidemiology 1986;123:577-591.

41. Fredman L, Schoenbach VJ, Kaplan BH, Blazer DG, James AS, Kleinbaum DG, et al. The Association between depressive symptoms and mortality among older participants in the Epidemiologic Catchment Area-Piedmont Health Survey. Journal of Gerontology 1989;44:S149-S156.

42. Pinsky JL, Jette AM, Branch LG, Kannel WB, Feinleib M. The Framingham Disability Study: relationship of various coronary heart disease manifestations to disability in older persons living in the community. American Journal of Public Health 1990;80:1363 1367.

43. Sahtahmasebi S, Davies R, Wenger GC. A Longitudinal analysis of factors related to survival in old age. Gerontologist 1992;32:404-413.

44. De Andino RM, Conde-Santiago JG, Mendoza MM. Functional disability and mental impairment as predictors of mortality in community-dwelling elderly Puerto Ricans. Puerto Rican Health Science Journal 1995;14:285-287.

45. Kivinen P, Halonen P, Eronen M, Nissinen A. Self-rated health, physician rated health and associated factors among elderly men: the Finnish cohorts of Seven Countries Study. Age Ageing 1998;27:41-47.

46. Woo J, Ho SC, Yuen YK, Yu LM, Lau J. Cardiovascular risk factors and 18-month mortality and morbidity in an elderly Chinese population aged 70 year and over. Gerontology 1998;44:51-55.

47. Wallace RB, Woolson RF. The Epidemiologic study of the elderly. Oxford: Oxford University Press; 1992.

48. Nelson LM, Longstreth WT, Koepsell TD, Van Belle G. Proxy respondents in epidemiologic research. Epidemiologic Review 1990;12:71-86.

49. Ministério da Saúde. Conselho Nacional de Saúde. Resolução 196/96. Diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Rio de Janeiro: Fiocruz; 1998.