Overcoming Challenges with Rural-Related Health Data: Lessons Learned from Published Studies

Casey P. Balio, PhD, Margaret M. Francisco, DrPH, MPH, Nicole Galler, DrPH, MPH, Olivia A. Sullivan, DrPH, MPH, EMT, Kate E. Beatty, PhD, MPH, Michael B. Meit, MA, MPH, Chyanne Morrison, MPH, Christiana Keinath, MSLIS, Stephanie M. Mathis, DrPH, MPH

INTRODUCTION

Notable health differences are evident across rural populations when compared to non-rural populations. Rural populations experience higher mortality rates due to multiple diseases and conditions, higher rates of disability, and limited access to important health services.[1] While our understanding of the health of rural communities has grown over time, research that continues to examine the unique needs of rural communities remains important for identifying opportunities to advance rural health.[2,3]

Rural health research is not without challenges and potential barriers, including those stemming from data quality, data availability, data timeliness, and data suppression.[4–6] Additionally, changing or differing definitions of rurality can pose challenges for understanding rural health needs and changes over time.[7–11] Many secondary datasets are available to assess community characteristics and health outcomes at state and national levels. While these datasets are useful for understanding a variety of health issues such as health care access, morbidity, and mortality, data can often be limited when focusing on rural areas.[4,5,12] Frequently cited challenges include delays in reporting, statistical and other biases, small samples, and lack of resources for data collection, among others.[4,5,12] For example, some national datasets generated through surveys, such as the Behavioral Risk Factor Surveillance System (BRFSS), may not sufficiently sample rural populations to provide statistical significance on topics of interest,[6] particularly at more granular geographic levels. This sampling limitation may be particularly impactful when seeking to examine the health needs, outcomes, and differences of rural subgroups. In order to conduct research that accurately reflects rural populations, it is imperative to

Key Takeaways

- Over half of included rural-related studies identified at least one potential strategy for overcoming study-related limitations or challenges, including those concerning data, within the limitations section.
- **Data source strategies** included using large, representative data sources and using established definitions of rurality.
- **Data analysis strategies** included using specific techniques like sensitivity analysis and testing multiple measures of rurality.
- Given the importance of rural data for understanding and improving the health of rural communities, researchers, data owners, and others should explore and use methods to facilitate robust and timely rural analyses.

understand data limitations, especially in secondary data sources, and how these can be taken into account, if not overcome. In the rural health literature, addressing limitations and mitigating data challenges is commonly focused on data collection methods, adding qualitative data, or use of local-level data generated by community organizations and local health departments.[6,13,14] While there have been calls for action to strengthen rural health research by increasing data availability, more information on methods for overcoming challenges related to the use of secondary data sources in rural health applications is needed.[4,5,14]

The purpose of this study is to identify potential strategies for overcoming study-related limitations or challenges, including those concerning data, within rural-related studies. Strategies as described in published studies that used secondary, quantitative data sources were of interest. These data sources are commonly leveraged for rural health initiatives and can carry a unique set of limitations relative to studies using primary data sources. This study involved systematically reviewing rural-related studies from reputable peer-reviewed journals that were published between 2019-2022. Findings provide insight into potential strategies for overcoming common challenges in rural health research. Policymakers, researchers, and others involved in collecting, analyzing, or reporting on rural health data could use them to identify opportunities for improvement.

METHODS

Study Design

A systematic review of peer-reviewed literature involving rurality was conducted to identify strategies that could help overcome challenges related to rural data. It was informed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).[15] The East Tennessee State University Institutional Review Board deemed this study non-human subjects research. A more detailed description of study methods is available elsewhere.[16]

Eligibility Criteria

Key criteria were selected to guide the inclusion of relevant studies. The review included studies that were published in English between 2019-2022 in select peer-reviewed journals: *American Journal of Public Health*; *Health Affairs; Health Services Research*; *JAMA*; or *Journal of Rural Health*. These five journals were prioritized for their track record of publications involving rural communities, health focus, and high impact factors. Studies needed to include a rural focus, be original research, and be set within the U.S. Additionally, studies needed to include quantitative data from a secondary data source and have a quantifiable, health-related outcome. Finally, studies needed to include an identifiable section with limitations as reported by the authors.

Search Strategy

In collaboration with a health sciences librarian, a search strategy was created. It included: 1) rural-related key words in the title or abstract (e.g., rural, non-metropolitan, and frontier); 2) journals of interest; and 3) article types for exclusion that were not relevant. Filters for year and language were incorporated. Using PubMed, the librarian completed the main search in February 2023, followed by a secondary search in March 2024 for a fifth journal.

Screening for Inclusion

Eligible studies were identified through a team-based screening process involving two stages. For the first stage, reviewers screened study titles and abstracts using an abbreviated version of the full eligibility criteria. Studies were randomly assigned to two teams. Each team consisted of two reviewers who independently screened their assigned studies and discussed differences to attain consensus. For the second stage, reviewers independently screened study full texts not excluded during the first stage using the full eligibility criteria. A second reviewer supported a final decision if a determination was unclear for a given study. As part of both stages, an initial pilot involving a random sample of studies was conducted, with any differences and revisions to the eligibility criteria discussed as a full team to attain consensus.

Data Extraction

Data elements were extracted from included studies through a structured, team-based process supported by Research Electronic Data Capture (REDCap).[17,18] First, reviewers completed two pilot reviews that each contained a random sample of 10 studies. Reviewers independently completed data extraction, with differences discussed as a team to attain consensus. These pilot reviews also supported process revisions for greater consistency. Next, the remaining studies were randomly assigned to three teams. Each team consisted of two reviewers who independently conducted data extraction for their assigned studies and discussed differences to attain consensus. A third reviewer examined any data elements in question. Lastly, preliminary results were collectively reviewed and refined as needed.

Included studies were examined for multiple data elements. Data elements of interest included study, data, and geographic characteristics as well as potential strategies for overcoming limitations or challenges as reported by the authors. Specifically, key characteristics included: topic area; study scope; geographic focus; types of secondary data sources; and measures of rurality. Types of potential strategies as described by study authors within an identifiable section on limitations were also extracted.

Data Analysis

Key characteristics of, and author-described strategies from the 259 included studies were examined using descriptive statistics. For many characteristics, categories were not mutually exclusive. Sums could exceed 100% or 259.

Limitations

Although this study offers actionable insight into potential strategies for overcoming challenges in rural health research, it is not without limitations. First, this study only captures strategies that were found within a publication section on author-reported limitations. Identified strategies were also not always directly tied to author-reported limitations (e.g., this strategy addressed this challenge or limitation). Second, included studies did not always concentrate on rural health, though they all involved rurality and health-related outcomes. Similarly, identified strategies did not always have a specific rural focus or application. Because of this broad inclusivity, findings are unlikely to reflect all strategies that could be leveraged to overcome challenges in rural health research. Third, this study was limited to studies using quantitative, secondary data published in select journals. This approach excludes, for example, studies that focus on primary data with qualitative methods and those that were published outside of those journals. Some of the limitations described herein may be similar across other types of data sources,

while others may differ or be unique to specific data types. Fourth, while this study purposefully integrated procedures to enhance rigor (e.g., double review), the potential for errors in data extraction and classification cannot be ruled out. Finally, the COVID-19 pandemic and the 2020 U.S. Census and subsequent data releases occurred during the time period for included studies (2019-2022), both of which could have implications for included studies and limitations or strategies cited within them.

RESULTS

Key Characteristics of Included Studies

The PubMed search generated 1,170 records for consideration. A total of 259 full-text articles ultimately met the inclusion criteria for this review. A diagram outlining the flow of records through the screening process, including reasons for exclusion, is available elsewhere.[16]

Key study, data, and geographic characteristics of included studies are shown in Table 1. A more detailed account of these characteristics is available elsewhere.[16] The studies spanned multiple topic areas, with the top three topics being health care workforce or availability (n=78, 30.1%), mental health or behavioral health (n=49, 18.9%), and social determinants of health (upstream drivers) (n=36, 13.9%). The majority of studies included a geographic focus on both rural and urban communities (n=233, 90%). Further, most studies were national in scope (n=187, 72.2%).

Many secondary data sources were used across studies. The three most commonly used types were administrative or medical records (n=137, 52.9%), survey (n=129, 49.8%), and claims (n=42, 16.2%). Similarly, the measure(s) of rurality varied across studies, with the most common being Rural-Urban Commuting Area (RUCA) Codes (n=62, 23.9%).

Table 1: Select Characteristics of Included Studies

Study Characteristic	Count of Studies [a]
Main Topic (Top Three) [b]	N (%)
Health care workforce or availability	78 (30.1)
Mental or behavioral health (including substance use)	49 (18.9)
Social determinants of health (upstream drivers)	36 (13.9)
Geographic Focus [c]	
Rural and urban	233 (90.0)
Rural or non-metropolitan only	26 (10.0)
Scope	
National	187 (72.2)
Single state	46 (17.8)
Subset of states	20 (7.7)
Sub-state region	6 (2.3)
Secondary Data Source (Top Three)	
Administrative or medical records	137 (52.9)
Survey	129 (49.8)
Claims	42 (16.2)
Measure of Rurality (Top Three)	
Rural-Urban Commuting Area (RUCA) Codes	62 (23.9)
Rural-Urban Continuum Codes (RUCC)	58 (22.4)
U.S. Office of Management and Budget (OMB)-based definition	37 (14.3)

[a] Some categories for study characteristics were not mutually exclusive. Totals may not sum to 100% or 259.
[b] A maximum of three primary topics were identified per included study.
[c] These categories were designed to include similar terminology.

Author-Described Strategies

Studies described a variety of strategies within the context of limitations sections that could help overcome study-related limitations or challenges. Over half of included studies identified at least one potential strategy (n=156, 60.2%). Among studies that described strategies, the most common types included data analysis strategies (n=99, 63.5%) and data source strategies (n=88, 56.4%). Sampling strategies (n=7, 4.5%) and other types of strategies (n=3, 1.9%) were described by a relatively smaller

proportion of studies. Within each of these types, authors described a variety of specific strategies. Examples of the three most common types of strategies are presented in Table 2.

Table 2: Examples of Common Types of Author-Described Strategies

Strategy Type	Examples
Data Analysis Strategies	Using techniques such as imputation to estimate missing or excluded values
	Conducting sensitivity or related analyses when working with small sample sizes or to account for changes over time
	Testing or using multiple measures for rurality or geography
	Collapsing categories of rurality
	Conducting analyses at the appropriate geographic unit based on available data or when describing potential uses for results (e.g., county-level versus ZIP code-level)
Data Source Strategies	Selecting, excluding, or combining data sources based on the level of available information and applicability to a study
	Using data sources that are credible, large, and representative of the population of interest
	Using data sources with current and accurate patient or facility location information, including at the ZIP code-level
	Using a measure of rurality that supports more granular levels of analysis (e.g., census tract)
	Using a common or established measure of rurality
Sampling Strategies	Using data sources that sample rural subgroups (e.g., Surveillance, Epidemiology, and End Results [SEER] data)
	Applying exclusion criteria to ensure equal measurement periods or reduce potential measurement error
	Limiting a sample based on specific characteristics (e.g., health conditions) to mitigate confounding

DISCUSSION

Research is pivotal to understanding the health needs of rural communities and potential programs and policies to address them. Challenges with underlying data and other limitations, however, can undermine the completeness and utility of rural health studies. As reported in published rural-related studies, this review describes common types of strategies that could be used to overcome such limitations. A majority of included studies identified at least one potential strategy in the limitations section, with a variety of data analyses, data sources, sampling, and other strategies described across studies. Overall, the findings offer lessons learned from the peer-reviewed literature that could inform future research and practice efforts focused on rural populations.

As part of the limitations section of included studies, the types of strategies most commonly described by authors focused on data sources and analyses. Rural or geographic considerations were present within both types. Data source strategies ranged from using large, representative data sources to using established definitions of rurality. Similarly, data analysis strategies ranged from using techniques like imputation or sensitivity analyses to testing multiple measures of rurality. The observed prevalence of data sources and analysis strategies could be related to the design of this study as studies needed to use

quantitative, secondary data, among other criteria, for inclusion. The prevalence of these strategies, however, underscores the importance of both carefully selecting appropriate data sources and applying appropriate statistical techniques when conducting rural health research involving secondary data. While these may represent important strategies overall, careful consideration should still be given to the potential implications (positive and negative) of applying a specific strategy within the context of a given study.

Considerations for Rural Health Research

- **Consider strengths and limitations of potential secondary data source(s).** A commonly reported strategy focused on using credible, large, and representative data sources. Researchers could carefully explore any potential secondary data sources that may be relevant to their study goals and population(s) of interest. A comprehensive understanding of underlying data collection methods, strengths, and limitations could support the selection and use of appropriate data sources.

- **Ensure the use of appropriate data analysis techniques.** Consistent with studies using secondary data, data analysis strategies were most commonly described. Researchers could carefully devise a statistical analysis plan, integrating techniques to address common limitations (e.g., missing values) where appropriate. Where feasible, consultation with researchers with relevant statistical expertise may be useful.

- **Describe any strategies used to overcome data or other study challenges.** Multiple studies did not describe even one potential strategy, at least in the limitations section. While potential strategies may have been described in other sections (e.g., methods), researchers should clearly discuss any key strategies used (or not used), including a rationale for their use (or lack of use) where possible. Such documentation could advance broader understanding and use of relevant strategies for overcoming challenges in rural health research.

- **Future research could leverage the rural voice.** People living in rural communities can provide valuable insight when designing studies, interpreting findings, identifying implications, and perhaps addressing select challenges in rural health research in ways that reflect rural contexts. Researchers could explore opportunities to include their perspectives during the research process. Ultimately, this inclusion may support the development of health-related interventions and policies that are not only evidence-based, but also culturally relevant and sustainable in rural contexts.

- **Future research should extend the findings of this study.** Overall, this study represents an important step in understanding common challenges in rural health research and potential strategies for overcoming them. Future research could continue to document these challenges and strategies. Examples of potential directions include describing strategies found in any section of published studies, reviewing published studies from other journals, and gathering insights directly from rural health researchers who frequently use secondary data sources.

Policy Opportunities for Strengthening Rural Health Data and Analysis

Building on study findings and previous work where appropriate,[4,5,8,13,14,16,19–22] including a prior policy brief from this study, opportunities that could strengthen rural health data collection, availability, and analyses are described.

- Assessing the feasibility of creating and making available datasets that are limited to samples of rural populations in order to support rural analyses, particularly for commonly used datasets with large sample sizes.

- Encouraging data owners to develop clear data documentation for data users that describes the use of any sampling techniques (e.g., oversampling) for rural populations.

- Providing resources for data owners (or others with appropriate expertise) to offer statistical consultation or technical assistance to data users aimed at promoting the appropriate use of commonly used datasets in rural analyses.

This project was supported by the Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (HHS) under the grant number U1CRH39978 Rural Health Research Grant Cooperative Agreement. The information or content and conclusions are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS, or the U.S. Government.

ETSU/NORC RURAL HEALTH RESEARCH CENTER

REFERENCES

1. Rural Health Disparities Overview. Rural Health Information Hub. Accessed May 8, 2024. https://www.ruralhealthinfo.org/topics/rural-health-disparities

2. Lutz A. The power of rural health research. National Rural Health Association. May 24, 2023. Accessed July 26, 2024. https://www.ruralhealth.us/blogs/2023/05/the-power-of-rural-health-research

3. Rural Health Information Hub. Healthcare Access in Rural Communities Overview. February 21, 2025. Accessed March 4, 2025. https://www.ruralhealthinfo.org/topics/healthcare-access

4. Scally CP, Burnstein E. *Rural Communities Need Better Data*. Urban Institute; 2020. Accessed January 22, 2025. https://www.urban.org/urban-wire/rural-communities-need-better-data

5. Scally C, Burnstein E, Gerken M, Immonen E. In Search of "Good" Rural Data. Urban Institute. Published online April 2020. https://www.urban.org/sites/default/files/publication/102134/in-search-of-good-rural-data.pdf

6. Zahnd WE, Askelson N, Vanderpool RC, et al. Challenges of using nationally representative, population-based surveys to assess rural cancer disparities. *Prev Med*. 2019;129 Suppl:105812. doi:10.1016/j.ypmed.2019.105812

7. Health Resources & Services Administration. How We Define Rural. Accessed December 4, 2024. https://www.hrsa.gov/rural-health/about-us/what-is-rural

8. Bennett KJ, Borders TF, Holmes GM, Kozhimannil KB, Ziller E. What is rural? Challenges and implications of definitions that inadequately encompass rural people and places. *Health Aff (Millwood)*. 2019;38(12):1985-1992. doi:10.1377/hlthaff.2019.00910

9. Long JC, Delamater PL, Holmes GM. Which definition of rurality should I use?: the relative performance of 8 federal rural definitions in identifying rural-urban disparities. *Med Care*. 2021;59:S413. doi:10.1097/MLR.0000000000001612

10. Hart LG, Larson EH, Lishner DM. Rural definitions for health policy and research. *Am J Public Health*. 2005;95(7):1149-1155. doi:10.2105/AJPH.2004.042432

11. Danek R, Blackburn J, Greene M, Mazurenko O, Menachemi N. Measuring rurality in health services research: a scoping review. *BMC Health Serv Res*. 2022;22:1340. doi:10.1186/s12913-022-08678-9

12. U.S. Census Bureau. Understanding and Using American Community Survey Data: What Users of Data for Rural Areas Need to Know. Published online October 2020. https://www.census.gov/programs-surveys/acs/library/handbooks/rural.html

13. Bekemeier B, Heitkemper E, Backonja U, et al. Rural public health data challenges during the COVID-19 pandemic: the case for building better systems ahead of a public health crisis. *J Public Health Manag Pract*. 2023;29(4):496-502. doi:10.1097/PHH.0000000000001726

14. Strong D, Del Grosso P, Burwick A, Jethwani V, Ponza M. Rural Research Needs and Data Sources for Selected Human Services Topics. U.S. Office of the Assistant Secretary for Planning and Evaluation.

Published online August 2005. https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//65071/rs_1.pdf

15. PRISMA. Welcome to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) website. Accessed August 9, 2024. https://www.prisma-statement.org

16. Mathis SM, Sullivan OA, Francisco MM, Galler N, Beatty K, Meit MB, Morrison C, Keinath C, Balio CP. A review of common limitations in rural-related studies in the peer-reviewed literature. Published online September 2025. https://www.etsu.edu/cph/rhrc/documents/rhrc_091025.pdf

17. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010

18. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208. doi:10.1016/j.jbi.2019.103208

19. Bekemeier B, Park S, Backonja U, Ornelas I, Turner AM. Data, capacity-building, and training needs to address rural health inequities in the Northwest United States: a qualitative study. *J Am Med Inform Assoc JAMIA*. 2019;26(8-9):825-834. doi:10.1093/jamia/ocz037

20. Bouchard L, Garcia I, Warrell J. Urban bias in rural data sets. Published online February 2023.

21. Zahnd W. The negative effect of differential privacy and synthetic data from the U.S. Census Bureau on rural health and communities. Published online September 2022.

22. Chiolero A, Tancredi S, Ioannidis JPA. Slow data public health. *Eur J Epidemiol*. 2023;38(12):1219-1225. doi:10.1007/s10654-023-01049-6