

# Big data in dermatology: Publicly available health care databases for population health research



Raghav Tripathi, MPH,<sup>a,b</sup> Rishabh S. Mazmudar, BS,<sup>a,b</sup> Konrad D. Knusel, MS,<sup>a,b</sup>  
Jeremy S. Bordeaux, MD, MPH,<sup>a,b</sup> and Jeffrey F. Scott, MD<sup>a,c</sup>  
Cleveland, Ohio, and Baltimore, Maryland

Health care databases are used in population health research to gain insights into diseases, treatments, outcomes, and disparities.<sup>1</sup> Using them to answer hypothesis-driven research questions can be difficult. We aimed to describe best practices for navigating publicly available health care databases in dermatology. We considered “publicly available” to include databases that any credentialed researcher can access with or without an associated fee. We focused on national databases in the United States. State-specific databases and databases available only to institutional members or by special permission were not included. We also included a few examples of the most prominent international databases for context. Characteristics of publicly available databases commonly used in dermatology and representative studies published in the *Journal of the American Academy of Dermatology* are provided in Table I.

## STUDY DESIGN

The first step when embarking on a population health research project using a publicly available health care database is to select the most appropriate database to answer a hypothesis-driven research question. A number of factors influence the choice of database, including the study population, exposure and predictor variables, primary outcomes, and analytic strategy (Table II). A statistician should be involved throughout all stages of research to avoid misrepresentation or inappropriate analysis of data. A thorough statistical plan is critical to avoid “fishing” for statistically significant results. Second, the specific population represented by a data set should be carefully noted to define the scope of the research

question, develop appropriate conclusions, and identify avenues for future research. For example, many data sets incorporate weighting and stratification schema to account for the complex sampling procedures required to represent a broader population. The analytic strategy should also consider the effect of changes in sampling methods over time on the representation of the data set.

## DATA ANALYSIS

When data analysis is approached, systematic data loss caused by the sampling procedures used should be evaluated. Although missing and mis-coded data do not necessarily invalidate analyses, researchers should consider the extent of data completeness and pattern of data skew when interpreting results. It is common to use a proxy variable when the desired data points do not exist (eg, receipt of different treatments as a proxy for disease severity).<sup>1</sup> When possible, proxy variables for particular diagnoses should be validated to help establish the proxy’s accuracy in representing the intended outcome. Similarly, outcomes and exposures should be validated when possible to precisely interpret data and draw appropriate conclusions.

Moreover, understanding the performance metrics and accuracy of variable definitions is important in ensuring that conclusions are accurate and justified. When applicable, the sensitivity, specificity, positive predictive value, and negative predictive value of codes and algorithms should be determined in lieu of manual review of health records (which is not possible with publicly available health care databases).<sup>32</sup> In particular, database studies often present incomplete information on confounding

From Case Western Reserve University School of Medicine, Department of Dermatology, Cleveland<sup>a</sup>; University Hospitals Cleveland Medical Center, Department of Dermatology<sup>b</sup>; and Johns Hopkins University School of Medicine, Department of Dermatology, Baltimore.<sup>c</sup>

Funding sources: None.

Conflicts of interest: None disclosed.

Reprints not available from the author(s).

Correspondence to: Raghav Tripathi, MPH, Department of Dermatology, University Hospitals Cleveland Medical Center, Lakeside 3500, 11100 Euclid Ave, Cleveland, OH 44106. E-mail: [raghav.tripathi@case.edu](mailto:raghav.tripathi@case.edu).

J Am Acad Dermatol 2020;83:1546-56.

0190-9622/\$36.00

© 2020 by the American Academy of Dermatology, Inc.

<https://doi.org/10.1016/j.jaad.2020.04.145>

**Table I.** Characteristics of publicly available health care databases commonly used in dermatology<sup>1-5</sup>

Database curator*	Data set	Available years	Sampling method	Population represented	Structure of data†	Prerequisites for access‡	Cost§	Approximate size	Study examples in JAAD
Healthcare Cost and Utilization Project	NIS, KID	1988–present (NIS), 1997–present (KID; every 3 y)	Stratified sample of national discharge data	Adult and pediatric inpatients	Identity variable: discharge code Sample outcomes: length of stay, cost of care	Purchase, online training	\$160 to \$1,000/y (student discount available)	7 million hospital stays each year; weighted, >20% of all inpatients	The national burden of inpatient dermatology in adults <sup>2</sup>
	National Emergency Department Sample	2006–present	Stratified sample of national ED discharge data	ED discharges	Identity variable: discharge code Sample outcomes: cost of care, disposition	Purchase, online training	\$500 to \$1,000/y (student discount available)	>30 million ED visits annually; weighted, >20% of all ED visits	Financial burden of emergency department visits for atopic dermatitis in the United States <sup>3</sup>
	National Readmissions Database	2010–present	Stratified sample of national discharge data for hospital readmissions	Hospital readmissions	Identity variable: discharge code Sample outcomes: readmission rates and costs	Purchase, online training	\$500 to \$1,000/y (student discount available)	18 million discharges annually; >58.2% of hospital stays	Dermatology-specific and all-cause 30-day and calendar-year readmissions and costs for dermatologic diseases from 2010 to 2014 <sup>4</sup>
National Cancer Institute	SEER	1973–present	Demographic and cancer data from national cancer registries (eg, tumor morphology, primary site, treatment)	Population-based cancer registries covering 34.6% of the US population	Identity variable: individual patient code Sample outcomes: survival, treatment modality	Online training	None	>10 million malignant and in situ cases	Incidence and survival of sebaceous carcinoma in the United States <sup>5</sup>
	Health Information National Trends Survey	2003–present; interrupted	Telephone survey of adults among the general population	Nationally representative	Identity variable: individual survey respondent code Sample outcomes: health behaviors, health literacy	None	None	≈3,000 to 6,000 participants annually	Rural-urban differences in behaviors to prevent skin cancer: an analysis of the Health Information National Trends Survey <sup>6</sup>

Continued

Table I. Cont'd

Database curator*	Data set	Available years	Sampling method	Population represented	Structure of data <sup>†</sup>	Prerequisites for access <sup>‡</sup>	Cost <sup>§</sup>	Approximate size	Study examples in JAAD
CMSII	Medicare Provider Utilization and Payment Data Part D (Prescriber)	2012–present	Physician-level prescriber billing data	Physicians serving patients with Medicare and Medicaid	Identity variable: NPI/name Sample outcomes: prescription rates and expenditure	None	None	All individual providers paid under Medicare programs	Trends in Medicare spending on topical immunomodulators and chemotherapies <sup>7</sup>
	Medicare Provider Utilization and Payment Data Part A (Inpatient)	2004–present	Medicare inpatient claims data	Inpatient Medicare claims	Identity variable: claims Sample outcomes: expenditures, procedures	None	None	All Medicare Part A claims	Biologic therapy adherence, discontinuation, switching, and restarting among patients with psoriasis in the US Medicare population <sup>8</sup>
	Medicare Provider Utilization and Payment Data Part B (Outpatient)	2012–present	Medicare outpatient claims data	Outpatient Medicare claims	Identity variable: claims Sample outcomes: expenditures, procedures	None	None	All Medicare Part B claims	Trends in phototherapy utilization among Medicare beneficiaries in the United States, 2000 to 2015 <sup>9</sup>
	SEER-Medicare	1991–present	Linked SEER and Medicare claims data of clinical, demographic, and death information for cancer patients	Population-based cancer registries covering 34.6% of the US population	Identity variable: individual patient code Sample outcomes: survival, stage, and diagnosis	Approved research proposal, purchase, online training	\$60 to \$260/y for each data file (entire database cannot be acquired)	>10 million malignant and in situ cases	Differences in melanoma outcomes among Hispanic Medicare enrollees <sup>10</sup>
	Open Payments	2013–present	Payments and transfers of value between manufacturers (drug/medical device companies) and physicians and teaching hospitals	All payments directly or indirectly to physicians and teaching hospitals	Identity variable: NPI/name Sample outcomes: fees/honoraria	None	None	All payments made to physicians and teaching hospitals (>11 million records, \$9 billion total)	Analysis of conflicts of interest in pharmaceutical payments made to Food and Drug Administration physician advisers after dermatologic drug approval <sup>11</sup>
	Medicaid Pharmacy Pricing Database	2012–present	Surveys of invoice prices for prescription	Nationally representative	Identity variable: individual drug	None	None	2,000–2,500 pharmacies	Pharmacy costs of specialty medications

			medications from retail community pharmacies		Sample outcomes: cost, accounting for manufacturer-pharmacy price reduction			sampled monthly	for plaque psoriasis in the United States <sup>12</sup>
Health and Human Services	Medical Expenditure Panel Survey	1996–present	Large-scale surveys for health care use habits	Nationally representative	Identity variable: individual patient Sample outcomes: health care use, expenditure	None	None	30,000 patients annually	Impact of the Patient Protection and Affordable Care Act on dermatologic health care use <sup>13</sup>
American College of Surgeons	NSQIP, NCDB	2005–present (NSQIP), 2004–present (NCDB)	Submitted cases from participating academic centers	Cases from specific hospitals	Identity variable: individual patient Sample outcomes: treatment modality, outcomes	Institutional subscription (NSQIP), approved research proposal (NCDB)	None	500,000 to 1 million cases annually (NSQIP), >70% of all newly diagnosed cancer cases (NCDB)	Pathologic nodal evaluation improves prognostic accuracy in Merkel cell carcinoma <sup>14</sup>
Centers for Disease Control and Prevention, National Center for Health Statistics	National Health and Nutrition Examination Survey	1999–present	Survey-weighted individual questionnaires	Nationally representative	Identity variable: individual survey respondent code Sample outcomes: health behavior, diagnoses	None	None	5,000 interviews/y	The prevalence of acne on the basis of physical examination <sup>15</sup>
	National Health Interview Survey	1957–present	Cross-sectional clustered household interview survey	Nationally representative	Identity variable: individual or household survey respondent code Sample outcomes: insurance coverage, health care use	None	None	35,000 households (≈ 87,500 unique participants)	Sunburn prevalence among US adults, National Health Interview Survey 2005, 2010, and 2015 <sup>16</sup>
	National Hospital Discharge Survey, National Hospital Care Survey	1965–2010	Combination of purchased electronic data files from hospitals and independent organizations and manual entry from hospital records	Nationally representative	Identity variable: discharge code Sample outcomes: admission rates, length of stay, mortality	None	None	Discharges from 239 hospitals annually (>270,000 inpatient stays)	Inpatient hospital care for psoriasis: a vanishing practice in the United States <sup>17</sup>

Continued

Table I. Cont'd

Database curator*	Data set	Available years	Sampling method	Population represented	Structure of data <sup>†</sup>	Prerequisites for access <sup>‡</sup>	Cost <sup>§</sup>	Approximate size	Study examples in JAAD
	National Ambulatory Medical Care Survey	1973–1981, 1985, 1989–present	Survey of physicians regarding 30 patient visits during a 1-wk period	Representative of the nation, 4 census regions, and 34 states	Identity variable: individual patient visit Sample outcomes: visit duration, diagnoses	None	None	>3,000 physicians at >104 sites annually	Increasing utilization of dermatologists by managed care: an analysis of the National Ambulatory Medical Care Survey, 1990–1994 <sup>18</sup>
	National Hospital Ambulatory Medical Care Survey	1992–present	4-wk interviews at facilities selected by systematic random sampling	Regionally representative	Identity variable: individual hospital stay Sample outcomes: hospitalization rates, treatments	None	None	≈ 550 hospitals annually	The burden of skin diseases: 2004 <sup>19</sup>
	Youth Risk Behavior Survey	1990–present	Surveys conducted at individual schools every 2 y	Representative of all public/private school students in grades 9–12	Identity variable: individual survey respondent code Sample outcomes: health beliefs and behavior	None	None	>4.4 million high school students in >1900 surveys	Prevalence of sunburn, sun protection, and indoor tanning behaviors among Americans: review from national surveys and case studies of 3 states <sup>20</sup>
	Behavioral Risk Factor Surveillance System	1984–present	Telephone surveys obtained monthly through random-digit dialing	Poststratification weighting scheme (originally designed to be representative of individual states)	Identity variable: individual survey respondent code Sample outcomes: health behavior, diagnoses	None	None	≈ 400,000 adult interviews annually	Alcohol consumption and self-reported sunburn: a cross-sectional, population-based survey <sup>21</sup>
Food and Drug Administration	FDA Adverse Event Reporting System	1968–present	All adverse events and medication error reports submitted to the FDA	Representative of only reported adverse events	Identity variable: individual adverse event report Sample outcomes: report rates, event details	None	None	>18 million total adverse event reports	An analysis of reports of depression and suicide in patients treated with isotretinoin <sup>22</sup>
National Heart, Lung, and Blood Institute	Women's Health Initiative	1993–2005	Women 50–79 y enrolled in randomized controlled trials for hormone	Nonrepresentative	Identity variable: individual patient code Sample outcomes: diagnoses,	Approved research proposal	None	>93,000 women (observational component), 68,132 women (randomized controlled)	Melanoma risk prediction using a multi-locus genetic risk score in the

IBM	Explorys Claims—EMR Data	1998—present	Multihealth system data/analytics of electronic medical records, claims systems, practice management systems, and laboratories	Representative of patients within a health care organization	therapy, dietary modification, or supplementation, or to an observational study	health behaviors		trial component)	Women's Health Initiative cohort <sup>23</sup>
	MarketScan Research Database (formerly Truven)	1995—present	Administrative claims for commercially insured adults and those with Medicare supplemental insurance from employers and patients with Medicaid in 11 states	Representative of the privately insured population	Identity variable: individual patient code Sample outcomes: diagnoses, treatment modalities	By consultation with IBM	>300,000 health care providers, >315 billion data records from 63 million patients	Prevalence estimates for chronic urticaria in the United States: a sex- and age-adjusted population analysis <sup>24</sup>	
Optum	Optum Clinformatics DataMart	2000—present	Administrative claims and laboratory data (ie, laboratory results) for privately insured patients and Medicare beneficiaries	Representative of privately insured population	Identity variable: individual patient code Sample outcomes: treatments, laboratory results	By consultation with Optum	>13 million unique individuals annually	The clinical utility of laboratory monitoring during isotretinoin therapy for acne and changes to monitoring practices over time <sup>26</sup>	
DVA	Veterans Hospital Patient Database	2014—present (varies)	Registry of veterans in the United States	Representative of veterans receiving health care at Veterans Affairs hospitals	Identity variable: individual veteran/visit code (varies) Sample outcomes: visit diagnoses, treatment modality	Guidelines specific to individual database None	Varies substantially by DVA database	Racial disparities in the impact of chronic pruritus: a cross-sectional study on quality of life and resource utilization in United States veterans <sup>27</sup>	

Continued

Table I. Cont'd

Database curator*	Data set	Available years	Sampling method	Population represented	Structure of data <sup>†</sup>	Prerequisites for access <sup>‡</sup>	Cost <sup>§</sup>	Approximate size	Study examples in JAAD
IMS	IMS LifeLink Health Plan Claims Database and Pharmetrics Plus	2006–present	Commercial database of longitudinal inpatient and outpatient medical and pharmaceutical claims data and enrollment information	Varying representation based on specific data set	Identity variable: individual patient code Sample outcomes: prescription details, diagnoses	Online training specific to data set	By consultation with IMS	Longitudinal data for fully adjudicated medical and pharmaceutical claims of >77 million patients with 75 health plans	The risk of malignancy among biologic-naïve pediatric psoriasis patients: a retrospective cohort study in a US claims database <sup>28</sup>
Examples of publicly available international data sources <sup>¶</sup>	THIN	1994–present	Longitudinal medical record data set in the UK with information from general practitioners regarding demographics, encounters, diagnoses, and prescriptions	Representative of England and Wales	Identity variable: individual patient code Sample outcomes: treatment details, visit diagnoses	Approved protocol, online training	By consultation with THIN	587 health care practices, >12 million patients (>3.6 million active patients) across the UK	Duration of oral tetracycline-class antibiotic therapy and use of topical retinoids for the treatment of acne among general practitioners (GP): a retrospective cohort study <sup>29</sup>
	Clinical Practice Research DataLink (eg, cancer data, admitted patient care data)	1987–present	Medical records from the National Health Service and private hospitals in the UK (diagnoses, symptoms, and prescription codes)	Representative of the UK population	Identity variable: individual patient code Sample outcomes: diagnoses, laboratory results	Approved research proposal, online training	£15,000 to £330,000, depending on study components (additional costs for data linkage)	>45 million patients (>13 million active patients) across the UK	Epidemiology of hyperhidrosis in 2 population-based health care databases <sup>30</sup>
	National Health Insurance Research Database	2000–present	Patient registration files and original claims data	Representative of Taiwan's population	Identity variable: individual patient code Sample outcomes: diagnoses, procedures, prescription details	Approved research proposal	\$15 per disc, \$6 per gigabyte of data	99.9% of Taiwan's population	Association between antidiabetic drugs and psoriasis risk in diabetic patients: results from a nationwide nested case-control study in Taiwan <sup>31</sup>

CMS, Centers for Medicare & Medicaid Services; DVA, Department of Veterans Affairs; ED, emergency department; FDA, Food and Drug Administration; KID, Kid's Inpatient Database; JAAD, *Journal of the American Academy of Dermatology*; NCDB, National Cancer Database; NIS, National Inpatient Sample; NPI, National Provider Identifier; NSQIP, National Surgical Quality Improvement Program; SEER, Surveillance, Epidemiology, and End Results Program; THIN, The Health Improvement Network; UK, United Kingdom.

\*All information is current as of December 1, 2019.

<sup>†</sup>Structure of data includes the specific identity variable that can be used to combine relational databases, as well as sample outcome variables for each database. This can help researchers evaluate whether these data sets can be linked and whether they contain outcomes relevant to one's study. For detailed information regarding which variables and data elements are included in each data set, visit the database website (see "Database Websites" footnote).

<sup>‡</sup>Prerequisites for access: most databases require a data use agreement. As such, data use agreement requirement was not listed in this column.

<sup>§</sup>Cost: Costs for databases can vary significantly over time, differ by accessing organization (eg, different prices for commercial and academic researchers, student discounts), and be dependent on the amount or type of data needed for the study. It is important to identify the specific costs relevant to one's study through the database curator.

<sup>||</sup>CMS: There are several data sets produced by CMS that are amalgamations/linked files of Medicare Parts A, B, and D data, all of which can be downloaded separately (eg, Health Care Information System, Healthcare Cost Report Information System, Medicare Enrollment Data, National Health Expenditure Data, Medicare Current Beneficiary Survey, Opioid Prescriber Data, Chronic Conditions Warehouse, containing Part A, B, and D). We have included only the core data files in this table to summarize which data are available from CMS. There is also an option to purchase Medicare claims data outright, which can be found on the CMS website. Visit the CMS website (see "Database Websites" footnote) for data downloads and additional information regarding how individual and compiled or linked data sets were produced.

<sup>¶</sup>Database websites, listed in order of presentation:

National Inpatient Sample: <https://www.hcup-us.ahrq.gov/nisoverview.jsp>

Kid's Inpatient Dataset: <https://www.hcup-us.ahrq.gov/kidoverview.jsp>

National Emergency Department Sample: <https://www.hcup-us.ahrq.gov/nedsoverview.jsp>

National Readmissions Database: <https://www.hcup-us.ahrq.gov/nrdoverview.jsp>

Surveillance, Epidemiology, and End Results: <https://seer.cancer.gov/>

Health Information National Trends Survey: <https://hints.cancer.gov/>

Medicare Provider Utilization and Payment Data Part A (Inpatient), B (Outpatient), and D (Provider): <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems>

SEER-Medicare: <https://healthcaredelivery.cancer.gov/seermedicare/>

Open Payments: <https://openpaymentsdata.cms.gov/>

Medicaid Pharmacy Pricing Database: <https://www.medicaid.gov/medicaid/prescription-drugs/pharmacy-pricing/index.html>

Medical Expenditure Panel Survey: <https://www.meps.ahrq.gov/mepsweb/>

National Surgical Quality Improvement Program: <https://www.facs.org/quality-programs/acs-nsqip>

National Cancer Database: <https://www.facs.org/quality-programs/cancer/ncdb>

National Health and Nutrition Examination Survey: <https://www.cdc.gov/nchs/nhanes/index.htm>

National Health Interview Survey: <https://www.cdc.gov/nchs/nhis/index.htm>

National Hospital Discharge Survey: <https://www.cdc.gov/nchs/nhds/index.htm>

National [Hospital] Ambulatory Medical Care Survey: <https://www.cdc.gov/nchs/ahcd/index.htm>

Youth Risk Behavior Survey: <https://www.cdc.gov/healthyyouth/data/yrbs/index.htm>

Behavioral Risk Factor Surveillance System: <https://www.cdc.gov/brfss/index.html>

FDA Adverse Event Reporting System: <https://open.fda.gov/data/faers/>

Women's Health Initiative: <https://www.whi.org/>

IBM Explorys: <https://www.ibm.com/watson-health/about/explorys>

Truven: <https://marketscan.truvenhealth.com/marketscanportal/>

Optum: <https://www.optum.com/solutions/life-sciences.html>

Department of Veterans Affairs Databases: [https://www.ea.oit.va.gov/eaoit/va\\_ea/opendata.asp](https://www.ea.oit.va.gov/eaoit/va_ea/opendata.asp)

Integrated Medical Systems Health: <https://www.iqvia.com/>

The Health Improvement Network: <https://www.the-health-improvement-network.com/>

Clinical Practice Research Datalink: <https://www.cprd.com/>

National Health Insurance Research Database: <https://nhird.nhri.org.tw/en/>



**Table II.** Choosing an appropriate publicly available health care database for population health research

Tasks	Factors to consider	Database examples
Establish your study budget	Although many publicly available data sets are free, some require a fee	Free (SEER) Subscription-based payment (SEER-Medicare) 1-time payment (HCUP data sets)
Determine your study population	Consider population-specific data sets and registries	Medicare patients (SEER-Medicare and other CMS data) Children (KID) Emergency department visits (NEDS) Cancer (SEER, NCDB)
Determine the inclusion criteria	Consider disease-specific data sets and registries If not patient-based, consider a data set that specifically evaluates the inclusion criteria of interest	Conflicts of interest (open payments) Drug pricing (Medicaid Pharmacy Pricing Database)
List the exposure and predictor variables for your question	Ensure the data set contains the exposure and predictor variables relevant to your question  If the data set does not include all exposure and predictor variables, determine whether an acceptable proxy variable exists	Sociodemographic data (most data sets) Stage of disease (SEER) Treatment information (SEER-Medicare) Quartile of estimated median household income of residents in the patient's zip code in lieu of income (HCUP data sets)
Determine whether the study is cross-sectional or longitudinal	Cross-sectional Longitudinal If the study is longitudinal but an appropriate data set does not exist, consider combining several years of a cross-sectional data set	HCUP data sets MEPS Trends during several years of NIS data (eg, 2002–2012)
List the primary outcomes	Consider a data set that contains the most granular details regarding the specific outcomes of interest	Cost of care (HCUP data sets) Length of hospital stay (NIS) Survival/mortality (SEER) Prescribing habits or procedure breakdown for health care providers (Medicare Part D) Healthcare use (MEPS)

CMS, Centers for Medicare & Medicaid Services; HCUP, Healthcare Cost and Utilization Project; KID, Kid's Inpatient Dataset; MEPS, Medical Expenditure Panel Survey; NCDB, National Cancer Database; NEDS, National Emergency Department Sample; NIS, National Inpatient Sample; SEER, Surveillance, Epidemiology, and End Results.

variables and qualitatively discuss the potential for confounding without a quantitative evaluation of the extent of bias. Assessing systematic bias through sensitivity analyses helps to quantify the residual confounding and clarify the implications for associations identified through statistical analyses.

Finally, consider linking multiple data sets when addressing complex research questions. For example, the relationship between regional socio-demographic data and various health outcomes can be analyzed by linking census data, area resource files, or geographic information system files with deidentified health care data.<sup>33</sup> Identifying a systematically collected, precise, and accurate variable in each data set (eg, zip code, National Provider Identifier) is required to link multiple data sets.

## RESULT REPORTING

Use of judicious language in reporting results (eg, correlation versus causation) is essential to ensure that conclusions are appropriately interpreted for clinical use and further prospective studies. Additionally, researchers should differentiate between statistical and clinical significance because analyzing large data sets commonly yields narrow confidence intervals and small *P* values. Carefully interpreting the clinical relevance of statistically significant odds ratios approaching 1.0 helps limit overemphasis of clinically insignificant findings.

Transparency is a critical component of publicly available health care database research. Collaborative research initiatives for observational studies, including Strengthening the Reporting of Observational Studies

in Epidemiology and Reporting of Studies Conducted Using Observational Routinely-Collected Data, are endorsed by several biomedical journals and provide checklists for use when reporting observational data.<sup>34,35</sup> Adhering to these guidelines helps promote the reliability and reproducibility of research using publicly available databases.

## CONCLUSION

With the increasing availability, size, and use of publicly available health care databases in dermatology, maintaining methodologic rigor throughout study design, data analysis, and result reporting is paramount. Implementing best practices ensures that researchers can take full advantage of the breadth and depth of these databases, report thorough and accurate results, and develop valid conclusions that will ultimately guide clinical practice, health care policy, and future research.

## REFERENCES

1. Takeshita J, Gelfand JM, Li P, et al. Psoriasis in the US Medicare population: prevalence, treatment, and factors associated with biologic use. *J Invest Dermatol*. 2015;135(12):2955-2963.
2. Arnold JD, Yoon SJ, Kirkorian AY. The national burden of inpatient dermatology in adults. *J Am Acad Dermatol*. 2019;80(2):425-432.
3. Kwa L, Silverberg JI. Financial burden of emergency department visits for atopic dermatitis in the United States. *J Am Acad Dermatol*. 2018;79(3):443-447.
4. Zhang M, Markova A, Harp J, Dusza S, Rosenbach M, Kaffenberger BH. Dermatology-specific and all-cause 30-day and calendar-year readmissions and costs for dermatologic diseases from 2010 to 2014. *J Am Acad Dermatol*. 2019;81(3):740-748.
5. Tripathi R, Chen Z, Li L, Bordeaux JS. Incidence and survival of sebaceous carcinoma in the United States. *J Am Acad Dermatol*. 2016;75(6):1210-1215.
6. Zahnd WE, Goldfarb J, Scaife SL, Francis ML. Rural-urban differences in behaviors to prevent skin cancer: an analysis of the Health Information National Trends Survey. *J Am Acad Dermatol*. 2010;62(6):950-956.
7. Song H, Adamson AS, Mostaghimi A. Trends in Medicare spending on topical immunomodulators and chemotherapies. *J Am Acad Dermatol*. 2018;78(1):173-175.
8. Doshi JA, Takeshita J, Pinto L, et al. Biologic therapy adherence, discontinuation, switching, and restarting among patients with psoriasis in the US Medicare population. *J Am Acad Dermatol*. 2016;74(6):1057-1065.e4.
9. Tan SY, Buzney E, Mostaghimi A. Trends in phototherapy utilization among Medicare beneficiaries in the United States, 2000 to 2015. *J Am Acad Dermatol*. 2018;79(4):672-679.
10. Rouhani P, Arheart KL, Kirsner RS. Differences in melanoma outcomes among Hispanic Medicare enrollees. *J Am Acad Dermatol*. 2010;62(5):768-776.
11. Kuschel SL, Ricotti CM, Dunnick CA, Hugh J, Dellavalle RP. Analysis of conflicts of interest in pharmaceutical payments made to Food and Drug Administration physician advisers after dermatologic drug approval. *J Am Acad Dermatol*. 2019;81(6):1419-1420.
12. Yang EJ, Beck KM, Sekhon S, Bhutani T. Pharmacy costs of specialty medications for plaque psoriasis in the United States. *J Am Acad Dermatol*. 2019;80(1):274-275.
13. Tripathi R, Knusel KD, Ezaldein HH, Bordeaux JS, Scott JF. Impact of the Patient Protection and Affordable Care Act on dermatologic health care utilization. *J Am Acad Dermatol*. 2019;81(2):631-634.
14. Lemos BD, Storer BE, Iyer JG, et al. Pathologic nodal evaluation improves prognostic accuracy in Merkel cell carcinoma: analysis of 5823 cases as the basis of the first consensus staging system. *J Am Acad Dermatol*. 2010;63(5):751-761.
15. Stern RS. The prevalence of acne on the basis of physical examination. *J Am Acad Dermatol*. 1992;26(6):931-935.
16. Holman DM, Ding H, Berkowitz Z, Hartman AM, Perna FM. Sunburn prevalence among US adults, National Health Interview Survey 2005, 2010, and 2015. *J Am Acad Dermatol*. 2019;80(3):817-820.
17. Stern RS, Bauer E, Epstein JH, et al. Inpatient hospital care for psoriasis: a vanishing practice in the United States. *J Am Acad Dermatol*. 2003;49(3):445-450.
18. Feldman SR, Fleischer J, Williford PM, White R, Byington R. Increasing utilization of dermatologists by managed care: an analysis of the National Ambulatory Medical Care Survey, 1990-1994. *J Am Acad Dermatol*. 1997;37(5 1):784-788.
19. Bickers DR, Lim HW, Margolis D, et al. The burden of skin diseases: 2004. A joint project of the American Academy of Dermatology Association and the Society for Investigative Dermatology. *J Am Acad Dermatol*. 2006;55(3):490-500.
20. Buller DB, Cokkinides V, Hall HI, et al. Prevalence of sunburn, sun protection, and indoor tanning behaviors among Americans: review from national surveys and case studies of 3 states. *J Am Acad Dermatol*. 2011;65(5 suppl. 1):S114.e1-S114.e11.
21. Mukamal KJ. Alcohol consumption and self-reported sunburn: a cross-sectional, population-based survey. *J Am Acad Dermatol*. 2006;55(4):584-589.
22. Wysowski DK, Pitts M, Beitz J. An analysis of reports of depression and suicide in patients treated with isotretinoin. *J Am Acad Dermatol*. 2001;45(4):515-519.
23. Cho HG, Ransohoff KJ, Yang L, et al. Melanoma risk prediction using a multilocus genetic risk score in the Women's Health Initiative cohort. *J Am Acad Dermatol*. 2018;79(1):36-41.e10.
24. Wertenteil S, Strunk A, Garg A. Prevalence estimates for chronic urticaria in the United States: a sex- and age-adjusted population analysis. *J Am Acad Dermatol*. 2019;81(1):152-156.
25. Shah K, Mellars L, Changolkar A, Feldman SR. Real-world burden of comorbidities in US patients with psoriasis. *J Am Acad Dermatol*. 2017;77(2):287-292.e4.
26. Barbieri JS, Shin DB, Wang S, Margolis DJ, Takeshita J. The clinical utility of laboratory monitoring during isotretinoin therapy for acne and changes to monitoring practices over time. *J Am Acad Dermatol*. 2020;82(1):72-79.
27. Shaw FM, Luk KMH, Chen KH, Wrenn G, Chen SC. Racial disparities in the impact of chronic pruritus: a cross-sectional study on quality of life and resource utilization in United States veterans. *J Am Acad Dermatol*. 2017;77(1):63-69.
28. Gu Y, Nordstrom BL. The risk of malignancy among biologic-naïve pediatric psoriasis patients: a retrospective cohort study in a US claims database. *J Am Acad Dermatol*. 2017;77(2):293-301.e1.
29. Barbieri JS, Hoffstad O, Margolis DJ. Duration of oral tetracycline-class antibiotic therapy and use of topical retinoids for the treatment of acne among general practitioners

- (GP): a retrospective cohort study. *J Am Acad Dermatol*. 2016;75(6):1142-1150.e1.
30. Ricchetti-Masterson K, Symons JM, Aldridge M, et al. Epidemiology of hyperhidrosis in 2 population-based health care databases. *J Am Acad Dermatol*. 2018;78(2):358-362.
  31. Wu CY, Shieh JJ, Shen JL, Liu YY, Chang YT, Chen YJ. Association between antidiabetic drugs and psoriasis risk in diabetic patients: results from a nationwide nested case-control study in Taiwan. *J Am Acad Dermatol*. 2015;72(1):123-130.
  32. Hsu DY, Dalal P, Sable KA, et al. Validation of *International Classification of Disease Ninth Revision* codes for atopic dermatitis. *Allergy*. 2017;72(7):1091-1095.
  33. Tripathi R, Knusel KD, Ezaldein HH, Scott JF, Bordeaux JS. Association of demographic and socioeconomic characteristics with differences in use of outpatient dermatology services in the United States. *JAMA Dermatol*. 2018;154(11):1286-1291.
  34. Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007;4(10):1628-1654.
  35. Benchimol EI, Smeeth L, Guttman A, et al. The Reporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.