Paradigm Project

AcademyHealth

Using Real-World Data and Artificial Intelligence to Advance Health Services Research

Summary

The field of health services research (HSR) can capitalize on burgeoning sources of real-world data to parse new and perennial questions about health care costs, quality, and access, as well as potentially increase the timeliness and relevance of research findings. But first, the field must prepare the infrastructure, including academia, research funding, and peer-reviewed publications, to deliver on the promise and avoid the pitfalls of greater and more sophisticated use of real-world data. While researchers have long used structured real-world data like claims to answer questions in policy and practice, myriad new unstructured data sources, including free text in electronic health records (EHRs) and images from X-rays and other technologies, are emerging for exploration. Similarly, new methodologies, such as machine learning and natural language processing, are being applied to real-world data to gain deeper insights about which care is right for which patients. This brief summarizes key points from a February 2021 meeting convened by AcademyHealth to examine greater use of real-world data in HSR and related issues, including safeguarding against the introduction of racial and other biases; addressing privacy concerns; establishing data standards; developing data resources as public goods; and helping researchers gain needed skills to design and conduct studies and interpret and disseminate findings.

Background

In recent years, advances in computing power have enabled researchers to leverage complex, large, and novel data sources to reveal new insights for decision-makers. While data science has spread into many sectors, HSR has been relatively slow to incorporate new data sources and analytics into the field's methodological toolbox.

The AcademyHealth Paradigm Project, supported by the Robert Wood Johnson Foundation, is a concerted, collaborative effort to increase the relevance, timeliness, quality, and impact of HSR through innovation.[1] AcademyHealth, through the Paradigm Project, convened a February 2021 meeting to explore using real-world data—also sometimes referred to as "big data"—and related artificial intelligence methods, such as machine learning and natural language processing, to enhance HSR capabilities to improve health and health care. Over the course of two afternoons, a group of health services researchers and data experts discussed how real-world data and evidence can complement traditional HSR approaches to identify and answer questions relevant to health policy and practice. Among the topics explored, participants discussed:

- Using nontraditional data sources and artificial intelligence methods alongside traditional HSR approaches to answer questions related to health care costs, quality, and access.

Genesis of this Brief:

This brief is based on a meeting of researchers and research users that took place virtually on February 24-25, 2021. AcademyHealth convened the meeting as part of its Paradigm Project, a concerted, collaborative effort to increase the relevance, timeliness, quality, and impact of health services research (HSR). Funded by the Robert Wood Johnson Foundation, the project is ideating and testing new ways to ensure HSR realizes its full potential to improve health and the delivery of health care. The Paradigm Project is designed to push HSR out of its comfort zone—to ask what works now, what doesn't, and what might work in the future. Additional information may be found on the project's website at https://academyhealth.org/ParadigmProject.

- Tapping real-world data sources to untangle the causal inference of policy and practice interventions on heterogeneous subgroups.
- Preparing the HSR infrastructure, including academia, research funding, and peer-reviewed publications, to capitalize on the promise and avoid the pitfalls of real-world data.
- Applying research findings to answer real-world policy and practice questions related to improving health and health care.

This brief summarizes the February meeting discussion, including using real-world data and analytics to move beyond average effects; safeguarding against the introduction of racial, ethnic, and other biases in real-world data analysis; addressing privacy concerns; establishing standards; developing data resources as public goods; and helping researchers gain needed skills to design and conduct studies and interpret and disseminate findings. The brief also examines the implications of greater use of real-world data for research funders, academia, peer-reviewed journals, and other aspects of the HSR ecosystem, including AcademyHealth. Because the session was off the record, the brief conveys the general content of the meeting without attributing specific comments to particular participants. The discussion was informed by existing research though neither the discussion nor this brief incorporates a systematic review of the literature related to using real-world data in HSR. A bibliography of some relevant, current literature is included at the end of the brief.

Real-World Data in Health Care

The U.S. Food and Drug Administration defines real-world data as "data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources."[2] Along with more traditional retrospective observational and survey data, real-world data can include information from EHRs, administrative and claims data, registries, patient-reported outcomes and wearable sensors, measures of social determinants of health, environmental exposures, and even clicks on a webpage, tweets, and geolocation data from smartphones and other mobile devices. Common users of real-world data and related research include pharmaceutical companies, payers and purchasers, providers, policymakers, and patients.[3]

In one example of how real-world data may increase the timeliness of HSR findings for policymakers, researchers are using geolocation data from 45 million smartphones to understand health care utilization and social mobility during the COVID-19 pandemic by examining people's visits to hospitals, physician offices, dental offices, and other health care sites. "This dataset has allowed us to look at these utilization patterns and see how they're changing in real time, such that policymakers may be able to make closer to real-time adjustments, whether in payment policy or trying to improve access," a participant said, adding that such analyses can help counter "policymakers routinely saying to health services researchers, we're too slow."

Volume. Velocity. Variety. Veracity. Value.

The so-called five Vs offer a helpful context to grasp the concept of big data.[4] As the volume, velocity, and variety of real-world data increase—fed by the exponential growth of digital information generated in the connected world we live in—ensuring the veracity and unlocking the value of real-world health data falls squarely in the HSR wheelhouse. While the field has long used structured real-world data like claims to answer research questions, myriad new unstructured data sources, including free text in EHRs and images from X-rays and other technologies, are emerging for exploration. Similarly, new methodologies, such as machine learning and natural language processing, are being applied to real-world data to gain deeper insights beyond traditional randomized controlled trials.

A form of artificial intelligence, machine learning essentially enables computers to learn and adapt by analyzing and drawing inferences from patterns in large datasets. Algorithms, or the step-by-step rules used in problem-solving calculations, are the fuel of machine learning. Similarly, natural language processing, or NLP, is another form of artificial intelligence that "helps computers understand, interpret and manipulate human language."[5] An offspring of linguistics, NLP enables computer software not only to read text and hear speech but interpret language, measure sentiment, and determine importance.

Moving Beyond Average Effects to Precision HSR

Increasingly, researchers are using real-world data and machine learning to move beyond average effects captured by randomized controlled trials. The goal is to pin down causal inference for subgroups that may respond differently to new drugs and medical devices post market, or perhaps even more importantly, identifying more generally across practice which tests and medical procedures work best for which patients, according to participants at the Paradigm Project gathering. Similarly, recent studies using machine learning to analyze large complex datasets have pinpointed patient-level differences related to physicians ordering low-value care and the impact of increased patient cost sharing for prescription drugs.

In the first study, researchers created algorithmic predictions of the results of testing patients in emergency departments for heart attacks. Using traditional analytic approaches, the testing on average appeared to be cost-effective. But the analysis powered by machine learning that accounted for more granular patient-level differences found that almost half of the tests should never have been ordered, and even more importantly, many patients who should have been

tested were not.[6] In the second study, researchers took advantage of a quirk in Medicare's prescription drug benefit structure to conduct a natural experiment.[7] At the time of the study, beneficiaries paid 25 percent out of pocket each year for prescription drugs until they reached $2,500 in spending, and then they paid 100 percent out of pocket for the next drug. Spending thresholds, however, were not pro-rated in beneficiaries' first calendar year of enrollment, and enrollment eligibility began in the month beneficiaries turned 65. So those born later in the year enrolled later in the year, and in turn had less time to reach the spending threshold, so they faced lower prices on average. Researchers trained an algorithm to identify patients who really needed certain drugs like statins and antihypertensives, finding that those exposed to higher cost sharing died at about a 30 percent higher rate than those who didn't face higher drug costs.

"We're used to thinking about averages…but patients are all different and machine learning and access to data are letting us do justice to those differences to where we can look for both high-value health care and low-value health care at the patient level," one participant said, adding, "So, it's precision health services research, if you want."

Automating Analysis Versus Generating Knowledge

More and more, machine learning and algorithms are being used to develop new clinical diagnostic tools, such as using artificial intelligence to interpret X-rays to diagnose knee pain or scan retinas for signs of diabetic retinopathy. But real-world data and machine learning in some cases can go beyond just reading an X-ray and generate new medical knowledge. In one recent study, for instance, researchers used machine learning to examine knee X-rays and linked the images to patient-reported pain symptoms. Not only could the algorithm do a better job than radiologists of explaining which patients felt pain, the algorithm also did a better job of explaining pain in Black patients, who historically have been undertreated for knee pain.[8]

"If we want algorithms to help make headway on understanding and producing medical knowledge, we can't just have them spit back out what a human would say about an image—that's good for health delivery purposes, where we want to make health care cheaper and more efficient and less error prone, but it's not going to get us far in building medical knowledge," a participant said.

Addressing Racial and Other Inequities

Despite their promise, machine learning and other artificial intelligence tools are not a "magic bullet," to solve health care cost, quality, and access problems, data experts at the meeting agreed, because algorithms can do enormous harm—even "automate errors"—if designed incorrectly. The conventional wisdom is that biased algorithms come from biased data, and biased data come from bias in society, and the only solution is to fix bias in society. Some participants, however, questioned that premise, pointing out that algorithms, instead of reinforcing structural disparities, can actually help dismantle disparities if constructed correctly.

For example, a 2019 study examined a commercial algorithm widely used to predict patients who would benefit from intensive care management interventions.[9] Researchers—who had access to the proprietary algorithm's inputs, outputs, and outcomes—found that the algorithm exhibited significant racial bias—not through ill intent, but through faulty use of past health care expenditures as a proxy for future health care needs.

At a given risk score, Black patients were considerably sicker than white patients—remedying the disparity increased the share of Black patients receiving additional help from 17.7 percent to 46.5 percent. Researchers reasoned that because Black people have unequal access to care and higher risk factors due to systemic racism, using their past health care costs to predict future health care needs introduced racial bias into the prediction. Such an approach is far from unique in the health sector where past claims data are relatively available and often used to predict future needs. Ultimately, researchers involved in the study retrained the algorithm using an index variable that combined cost prediction with health prediction, which reduced the racial bias substantially. "Just like any tool, [algorithms] can be a force for good, or they can be a force for evil, and which one it is, is kind of up to us when we build them," a participant concluded.

Unlike research to determine causal inference, such as differences in treatment effects among heterogeneous populations, research in a prediction world using machine learning is relatively straightforward. In a causal inference world, researchers' traditional focus on cleaning and correcting data grows exponentially because they might be working with 5 million variables in a complex dataset instead of five variables in a more traditional analysis.

"That's the bad news, the good news is that when we're working in a prediction world and not a causal inference world, we actually don't need to pay quite as much attention to all 5 million variables on the right-hand side of the model because all we want from those variables is their ability to predict what's on the left-hand side—what's the dependent variable that we're interested in," according to a researcher experienced in using real-world data and methods. "But the price of that is that we really need to make sure that that left-hand side variable is perfect, so…there's also a lot of very important work in making sure that the thing we're predicting is exactly what we think it is and that we're not using costs as a proxy for needs, because of the biases…because algorithms will key in on those differences and amplify them."

In discussing how to prevent bias from creeping into research using machine learning, meeting participants coined an acronym on the spot—GAP, for good algorithmic practice—as well as the need to establish standards for developing and using algorithms in HSR. Participants also discussed the need to strengthen the inter- and multi-disciplinary nature of HSR, pointing out the need for "bilingual" researchers conversant in health, economics, and data science, for example, who would spot the flaw in using past health care spending as a proxy for future needs of underserved patients.

Linking Diverse Data Sources at the Patient Level

Invoking the maxim that "all data are health data," participants stressed the importance of linking data sources at the patient level and getting data directly from patients through surveys, wearable medical devices, and sensors, including smartphones.

For example, linking real-time patient reports of how they are feeling in the moment and data from "wearables" to EHRs could give clinicians a fuller picture of health status and support shared decision making with patients. Potentially, clinicians could create near real-time feedback loops to engage patients through email, for instance, by prompting for patient-reported outcomes and replying with an intervention. "Using ecological momentary assessments—very frequent prompts to patients to report on their subjective status at that point in time—reduces recall bias and also generates a lot of data, and data that's very granular and has a lot of temporal richness to it," according to a physician researcher at the meeting.

Unlocking the power of patient-level data in both research and practice, however, is fraught with privacy concerns, making "patient trust" a key issue in accessing and linking data, one participant said. In contrast to the United States, other countries, particularly Denmark and Sweden, have "rich" datasets linking population health information at the individual level down to biomarkers. "They've somehow managed in multiple countries to be able to share national level population-based data at the individual level for researchers for essentially free," according to a researcher with knowledge of the Scandinavian datasets. But other participants questioned Americans' willingness to embrace such transparency of health information, with one saying, "I think there are a lot of people who are very envious of the Scandinavian countries' datasets, but we are never going to go there…. maybe someday… never is a long time."

Do You Know Where Your Data Are?

Privacy concerns are and will likely remain a major barrier to channeling the power of big data to inform health policy and practice, with one data expert saying, "If we're talking about using government capabilities in any way, privacy really has to be at the forefront of the protections that we're envisioning, as well as clearly delineating what the value proposition and potential benefits of any research application might be."

Several participants, however, noted that the private sector already collects and uses tremendous amounts of supposedly anonymous data from smartphones and other sources to analyze consumer behavior for advertising and other uses. But as The New York Times showed in 2018, it's relatively easy to connect that blue geolocation dot on your app screen to you and precisely track almost your every step.[10] In the case of industry, as one person said, "Frankly, they have a lot of incentive to not reveal the fact that they have access to this type of data, because it does look very creepy."

And while privacy concerns are real, at the same time, "there are huge risks to not having data," according to a participant who noted how the COVID-19 pandemic has highlighted the "shambles" of the nation's health data infrastructure to solve problems ranging from predicting the pandemic to coordinating hospital beds to distributing vaccines. Other participants stressed the need to create a "social license" for data access and raising public awareness and acceptance of using data for the "greater good," with one framing the issue as: "How do we get people to internalize that their data can be used for good purpose without it feeling like a threat."

One data expert discussed the idea of individuals donating their data, saying, "I'm fascinated by conversations about data ownership…and whether individuals could be compensated for the use of their data—whether they will eventually have some avatar working around them virtually that allows their data to be seeped out to some uses and blocked from others."

'Data is the New Oil'

Notwithstanding privacy concerns, the proliferation of data and the potential to monetize new insights into human activities and behavior have sparked comparisons of data as the new oil. Neither data nor oil has much value as a raw material—their value comes from refining and breaking down the parts and creating something new.[11] For health services researchers, accessing large real-world datasets can be expensive, literally millions of dollars. At the same time, not all data are equal, and quality is important, with one participant saying, "I always think of these black box data products that are on the market and being sold, and you don't have any insight as to how the data is being constructed, what's in there, where does it come from."

Given the potential to monetize data, competitive issues can prevent researchers from accessing datasets, and several participants cited the need to break down competitive barriers and develop data resources as "public goods" rather than commercial commodities. Building data repositories and reusing data could lower costs, and there is a pressing need as well to standardize data collection, for example, across health care payers and states. The glaring gaps in U.S. health data, for instance, are illustrated by missing race/ethnicity data for almost half of U.S. adults who had received at least one dose of a COIVD-19 vaccine by mid-March 2021.[12]

Several participants stressed the importance of developing "federated," or centralized, approaches to data collection and documentation. Under a federated model, multiple data sources feed into one another and are managed and documented in a standard fashion. Another pressing need is to develop best practices for data documentation, such as FAQs and publishing metadata—or data that describes and provides information about key aspects of other data. "We value data but we don't value the stewardship of that data," a participant observed.

Real-World Data and the HSR Ecosystem

As researchers increasingly embrace machine learning and real-world data, the surrounding ecosystem—research funders, academia, peer-reviewed journals, and the health care industry—also must evolve if the relevance, timeliness, quality, and impact of HSR is to increase. To some degree, the field, with the overt support of academia and the tacit support of funders, relies on investigator-initiated research that too often isn't sufficiently grounded in real-world problems, some participants observed.

"We have bought into the investigator-initiated model of health services research, because that is the coin of the realm in academic institutions," a participant said. "But we are an applied field… and I think we need to be responding to the priorities of policymakers, health system leaders, etc., and so I think that's a disconnect in the basic incentive structure."

For example, promotion and tenure policies in academia typically reward a track record of publishing in peer-reviewed journals not improving data linkages or devising better ways to document data. "For tenure, why do just publication's matter?" a participant asked. "Why can't data assets or radically improved linkage approaches or metadata … start counting, because they are just as important for knowledge building as the great publication using the data itself."

Similarly, researchers must forge new understandings with peer-reviewed journals, which will need to adopt data standards and identify qualified peer reviewers conversant in new methodologies. As one participant said, "In the near term, we're going to struggle, as we have these new methods, with dealing with reviewers in the journal space and making sure that they really understand how to interpret the work that we're going to submit and put out into the public space for interpretation and review."

The field's relationship with funders and industry—both within health care and beyond to the technology companies that collect and build large, novel datasets—also must change if HSR is going to use novel data to inform solutions to the real-world conundrums of a U.S. health care system that costs too much, harms too many patients, and leaves too many marginalized people without needed care. Funders, for example, could take a role in pushing for greater transparency and public engagement in plugging data gaps, such as the paucity of patient-level racial/ethnicity data, to inform policy and practice. Or they could partner with industry to purchase bulk access to data for research.

Unlike other fields such as computer science, where academics often have an entrepreneurial bent and form companies and partner with industry, the discipline of HSR hasn't really promoted itself to industry, according to a participant, who added that technology companies routinely recruit HSR graduate students with coding skills to work on projects. "Academia needs to figure out how to extract more value and how to partner better with industry [because industry] very much realizes that we need academia in order to do what we're doing right now," the participant said.

Others observed that health services researchers are skittish of commercial motives, with one recalling the field's existential crisis in the mid-1990s when the predecessor to the federal Agency for Healthcare Research and Quality drew the ire of surgeons and nearly was eliminated by Congress for overseeing guidelines questioning the appropriateness of spine surgery for uncomplicated low back pain.[13] "The value proposition for industry is only there as long as what you're showing is that you can sell more," the participant said, "and not there when you are trying to show that you can actually sell less… who's paying for research showing things shouldn't be done?"

Nonetheless, the field needs to identify ways to break down the "firewall" between researchers and industry because industry has the data researchers need. In the case of EHRs, vendors like Cerner and Epic "obviously have a commercial intent, but we have to get past that as researchers," one participant said, adding that EPIC founder Judy Faulkner created the EPIC Health Record Network, a public benefit corporation focused on research, because she doesn't "perceive that we want to partner."

Real-World Dissemination and Implementation

Similar firewall issues exist between researchers and the health care delivery system, with few researchers willing to get inside health systems and provider organizations to understand how they operate. Most academic researchers "come to health systems and say, 'Hey I just want your data,' you know kind of cut and run, and that doesn't help build meaningful lasting relationships where we're really trying to help these systems think differently and transform care," according to a participant.

Emerging models, such as embedding researchers in delivery systems, can help bridge the gap between research and practice, but "you can't just throw them into a health system with a dataset and say have fun." Researchers need to learn how health systems oper-

ate, and health system administrators need to learn how research works. For many health system administrators, according to a researcher experienced in working in a health system, "The idea of research is that you have the right answers that I need to implement, and once I implement it, there's going to be rainbows and puppies and everything is going to be perfect. And the reason we think that is because we're bombarded by vendors who say exactly that—that if you hire us, that if you use us, that if you purchase this, then everything will be perfect."

On the flip side, rather than diving into the data in search of a problem to solve, researchers need to understand what operational problems health systems are facing and then identify what data might help solve the problem. "Our frontline people have the questions—they know what they want to learn about. And their issue is they don't know how to get to the answer, so they'll say something like we have this new quality indicator on asthma, as it turns out we're bad at it, we'd like to be better at it. So, we're going to do X. Is this a good idea? And then the researcher shows up and says well you're collecting all the wrong data and there's no way we can answer this question for you."

Researchers and administrators sometimes speak different languages, the researcher continued, recounting a story about how his team curated data and rolled it out to frontline staff as a "self-service portal." The result: "Everybody hated us. I got hate mail because self-serve means that they're doing it." After consulting with the marketing department, researchers deployed the same system, calling it on-demand data, and "everybody loved it, and it was just this change in language."

On the policymaking front, similar nomenclature and culture differences can complicate communication of research findings. "We have a way of thinking about uncertainty and communicating it and talking about it and living with it, that is pretty different from the way people who have to actually make decisions do," a participant said. "We see this with COVID all the time, right? I mean what should the standard be for whether we recommend mask wearing or not. You can't say like, 'Well, we sort of think this and maybe we're going to learn more.' That doesn't seem to be a successful public relations strategy."

More broadly, there is a need to set up infrastructure to communicate findings. One model is the State-University Partnership Learning Network, managed by AcademyHealth to support evidence-based state health policy and practice with a focus on transforming Medicaid-based health care. For a variety of reasons, researchers often give "short shrift" to dissemination and communication of findings "often because they've run out of money, and dissemination costs money and getting people to pay attention to what you've learned… it really requires a different group of people to help you get that message out in a meaningful way."

Implications for HSR and AcademyHealth

As the data universe keeps expanding—from zettabytes to yottabytes and beyond—participants agreed that AcademyHealth can play an important role in helping the HSR field embrace new data sources and methods to identify solutions to real-world problems in policy and practice.

"I think AcademyHealth is well positioned to play this role in really helping to educate health services researchers," a participant said. "So, if you use me as sort of the average test case, I think there's a great lack of knowledge about nontraditional data sources—what's out there, how to use the data, how to get access to the data, what some of the analytic methods are in terms of analyzing big data and how to interpret the findings."

As the Paradigm Project continues to use human-centered design and other tools to innovate and identify ways to increase the relevance, timeliness, quality, and impact of HSR, integrating conversations about real-world data and community engagement and participation will be critical. Other areas where AcademyHealth can support the field in leveraging the use of real-world data include:

- Supporting standardization of data use and standards, including privacy protections, data documentation, model data use agreements, and other processes.
- Designing training and other educational programs to help researchers gain skills to use novel data and methods.
- Building relationships among health services researchers, academia, and industry—in both the technology sector and health care delivery.
- Working with funders and journals to align timescales to support near real-time publication of research findings based on real-time data that can support policy and practice.
- Building capacity to translate and communicate research results in accessible and actionable ways.
- Creating awards to recognize researchers using real-world data and methods to answer questions that inform policy and improve practice.

About the Author

Alwyn Cassil is a Principal at Policy Translation, LLC.

Endnotes

1. https://www.academyhealth.org/ParadigmProject.
2. Food and Drug Administration (FDA). Framework for FDA's Real-World Evidence Program. 2018. https://www.fda.gov/media/120060/download. Accessed March 10, 2021.
3. Rudrapatna VA, Butte AJ. Opportunities and challenges in using real-world data for health care. *J Clin Invest*. 2020;130(2):565-574. doi:10.1172/JCI129197.
4. Kalbandi, Ishwarappa & Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. Procedia Computer Science. 48. 319-324. 10.1016/j.procs.2015.04.188.
5. SAS. Natural Language Processing (NLP): What it is and why it matters. Accessed March 19, 2021, at https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html.
6. Mullainathan, S & Obermeyer, Z. Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care. National Bureau of Economic Research. Working Paper No. 26168. (2021). doi:10.3386/w26168. Accessed at https://www.nber.org/papers/w26168
7. Chandra, A, Flack, E & Obermeyer, Z. The Health Costs of Cost-Sharing. National Bureau of Economic Research. Working Paper No. 28439. (2021). Doi: 10.3386/w28439. Accessed at https://www.nber.org/papers/w28439
8. Pierson E, Cutler DM, Leskovec J, Mullainathan S, Obermeyer Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat Med*. 2021;27(1):136-140. doi:10.1038/s41591-020-01192-7
9. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/science.aax2342
10. Valentino-DeVries, J, Singer, N, Keller, MH, Krolik, A. The New York Times. Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret. (December 10, 2018). Accessed March 19, 2021, at https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.
11. The phrase, data is the new oil, dates to 2006 and is widely attributed to Clive Humby, a British mathematician.
12. Ndugga, N, et al. Latest Data on COVID-19 Vaccinations Race/Ethnicity. Kaiser Family Foundation. (2021) accessed March 25, 2021, at https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-race-ethnicity/.
13. Gray BH, Gusmano MK, Collins SR. AHCPR and the changing politics of health services research. Health Aff (Millwood). 2003;Suppl Web Exclusives:W3-307. doi:10.1377/hlthaff.w3.283

Appendix: Additional Reading

Abadie A. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. Journal of Eco Lit. Forthcoming.

Athey S. The Impact of Machine Learning on Economics. Unpublished Manuscript. 2018 Jan.

Athey S. Beyond prediction: Using big data for policy problems. Science. 2017 Feb 3; 355(6324):483-485.

Goldstein BJ, Rigdon J. Using Machine Learning to Identify Heterogeneous Effects in Randomized Clinical Trials – Moving Beyond the Forest Plot and into the Forest. JAMA Network Open Cardiology. 2019 March 8; 2(3).

Jarmin RS, O'Hara A. Big Data and the Transformation of Public Policy Analysis. JPAM. 2016 May 10; 35(3):715-721.

Jarmin RS, O'Hara A. Counterpoint to "Big Data for Public Policy: the Quadruple Helix. JPAM. 2016 May; 35(3):725-727.

Lane J. Big Data: The Role of Education and Training. JPAM. 2016 May 10; 35(3):722-724.

Mingle D. Healthcare: Moving Beyond Average. CIO Magazine. 2015 Oct 15.

Morgan FR, Wang D, Cebrian M, Rahwan I. The Evolution of Citation Graphs in Artificial Intelligence Research. Nature Machine Intelligence. 2019 Feb 11; 1:79–85.

McClelland R, Gault S. The Synthetic Control Methods as a Tool to Understand State Policy. The Urban Institute. 2017 March.

Mullainathan S, Obermeyer Z. On the Inequity of Predicting A While Hoping for B. AEA Papers and Proceedings. 2021 May. 111:37-42.

Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25; 36(6464):447-453.

Obermeyer Z, Emanuel EJ. Predicting the Future – Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016 Sept 29; 375:1216-1219.

Pavaloiu A, Kose U, Boz H. How to Apply Artificial Intelligence in Social Sciences. Unpublished Manuscript. 2017. Available at https://www.researchgate.net/profile/Hakan-Boz/publication/325398286_How_to_Apply_Artificial_Intelligence_in_Social_Sciences/links/5de3f853a6fdcc2837fd09eb/How-to-Apply-Artificial-Intelligence-in-Social-Sciences.pdf.

Schudde L. Heterogeneous Effects in Education: The Promise and Challenge of Incorporating Intersectionality Into Quantitative Methodological Approaches. Review of Research in Education. 2018 April 5; 42(1):72-92.

Sivarajah U, Kamal M, Irani Z, Weerakkody V. Critical analysis of Big Data challenges and analytical methods. J of Bus Res. 2016 Aug 10; 70:263-286.

Stetter C, Menning P, Sauer J. Going Beyond Average – Using Machine Learning to Evaluate the Effectiveness of Environmental Subsidies at Micro-Level. Contributed paper prepared for presentation at the 94th Annual Conference of the Agricultural Economics Society; 2020 April 15-17; KU Leuven, Belgium.

Thorlund K, Dron L, Park J, Mills E. Synthetic and External Controls in Clinical Trials – A Primer for Researchers. Clinical Epidemiology. 2020 May 8; 12:457-467.

Zou K, Li J, Imperato J, Potkar C, Sethi N, Edwards J, Ray A. Harnessing Real-World Data for Regulatory Use and Applying Innovative Applications. J Multidisciplinary Healthcare. 2020; 13:671-679.