

May 14, 2024

Administrator Chiquita Brooks-LaSure
Centers for Medicare and Medicaid Services

Re: Request for Information on Research Data Request and Access Policy Changes

Dear Administrator Brooks-LaSure:

Thank you for the opportunity to comment on recent proposals by the Centers for Medicare and Medicaid Services (CMS) to stop providing researchers with physical extracts of CMS data and instead provide data access solely through the Virtual Research Data Center (VRDC).¹ We share other researchers' concerns that, without improvements to the VRDC, this change could seriously impede research with CMS data, harming both CMS programs and the health care system overall.²

In our view, the VRDC's biggest shortcoming is its fee structure. At present, researchers receiving physical extracts pay a large upfront fee to acquire data but may reuse those extracts for a modest per-project fee. By contrast, researchers using the VRDC pay large recurring fees for each project and user seat. Thus, researchers who now use physical extracts will face a far higher marginal price of starting projects, adding team members, or investing more time in a project. They will likely also face at least somewhat higher total data access costs. These changes, especially the large increase in the marginal price of data access, will likely reduce how much research is produced.

In this letter, we make four recommendations for how CMS could mitigate the downsides of shifting all research into the VRDC if CMS decides to proceed with doing so:

- Offer institution- or funder-level "site licenses" to the VRDC: The VRDC's project and seat fees likely far exceed CMS' marginal costs of providing access. Allowing research institutions to purchase a "site license" under which they would pay project and seat fees closer to CMS' marginal costs could mitigate the large increase in the marginal price of data access associated with the transition to the VRDC but still allow CMS to recover its costs of providing data access. Ideally, such arrangements could even be implemented above the institutional level via negotiations with the relatively small number of federal and non-federal funders that support large amounts of research using CMS data.

¹ The views expressed in this letter are our own and do not necessarily reflect the views of the Brookings Institution or anyone affiliated with the Brookings Institution other than ourselves.

² See, for example, William L. Schpero, David J. Meyers, and Sarah H. Gordon, "Safeguarding Research Using Federal Health Insurance Data," *JAMA*, May 8, 2024, <https://doi.org/10.1001/jama.2024.5943>; Joshua Gottlieb and Kevin Rinz, "Proposed CMS Data Changes Risk Major Research Fallout," Briefing Book, March 4, 2024, <https://www.briefingbook.info/p/proposed-cms-data-changes-risk-major>; Academy Health, "RE: Research Data Request and Access Policy Changes, Announced February 12, 2024," February 26, 2024, https://academyhealth.org/sites/default/files/academyhealth_response_to_cms_on_data_access.pdf.

- Ensure that CMS' data access processes operate efficiently: While improving the *structure* of the VRDC's fees is important, the *level* of those fees matters too. Because those fees are set to cover CMS' costs, keeping fees affordable requires CMS to operate efficiently. Without appropriate cost discipline in CMS' operations, it will likely be difficult to fully (or even largely) mitigate the downsides of transitioning data access into the VRDC.
- Reconsider whether CMS should recover its full costs through fees: CMS currently recovers the full cost of providing data access through fees. Given the potentially large social benefits of research using CMS data, CMS should consider subsidizing these costs (especially the fixed costs of operating data access infrastructure), although we recognize that this would need to be weighed against other uses of CMS administrative funds.
- Set reduced fees for low-usage seats/projects: CMS should also set reduced fees for seats or projects with VRDC usage below a specified level. This would make it feasible for research team members who need to access raw data for a narrow set of purposes to do so, while also better accommodating projects that go through periods of inactivity.

The remainder of this letter examines these points in greater detail. Before proceeding, however, we note that the VRDC's fee structure is not the only downside to accessing data via the VRDC rather than physical extracts. For example, researchers using physical extracts can fully customize the hardware and software configuration of the computing facilities they use to analyze data, but that is not possible on the VRDC. Additionally, some types of data linkages are more difficult when using the VRDC. While our letter focuses on the VRDC's fee structure, CMS should also carefully consider how to mitigate these other downsides of migration to the VRDC.

Comparing CMS' fees for physical and virtual data access

CMS' fees for physical and virtual data access have very different structures. Researchers using physical extracts pay large upfront fees to acquire data; acquiring a full set of Medicare and Medicaid data for even a single year currently costs around \$200,000. But historically, after paying these fees, researchers (and others affiliated with the same institution) have been able to reuse those data as many times as they wish for a per-project fee of just \$2,000.

By contrast, researchers using the VRDC pay large recurring fees for each project and user seat (but do not pay upfront fees to acquire data). Thus, researchers using the VRDC face a far higher marginal price of starting new projects, adding team members, or investing more time in a project. Concretely, the price of a new VRDC project with a single user seat is at least \$35,000 for the first year and \$23,000 for each additional year, more than an order of magnitude higher than the price of a new project using physical extracts. And whereas allowing another team member to access CMS data involves no additional fees when using physical extracts, the VRDC charges fees for each additional seat of \$20,000 for the first year and \$13,000 for each additional year. Total costs are likely also higher in the VRDC for researchers who currently use physical extracts, particularly for large institutions with many active projects and users, albeit to a lesser degree.

A caveat is that the higher data access fees researchers incur in the VRDC may be offset to some degree by reduced costs of maintaining computing infrastructure. We doubt, however, that this would meaningfully offset the increase in the *marginal* price of an additional project, seat, or period of work, as the marginal cost of expanding computing infrastructure is typically relatively low. Moreover, such savings would often not accrue to individual researchers since computing infrastructure is often operated at the departmental or institutional level.

The higher fees in the VRDC setting—and especially the higher marginal price of data access under the VRDC fee structure—would reduce research output in a few important ways:

- Preventing use of CMS data by many types of researchers and for many types of projects. Some research projects are feasible only if data access fees are low. Some researchers, especially graduate students and other early career researchers, have little access to grant funding, and even well-funded researchers face limits on grant funding. Additionally, some types of high-value projects, particularly some types of exploratory or innovative research as well as quick-turnaround work that addresses timely policy questions, can be harder to fund via grants. Thus, many projects that are feasible under the fee structure that applies to physical extracts would not be feasible under the VRDC fee structure.
- Preventing efficient division of labor among research team members. Conducting research efficiently often involves assigning different roles to different team members. For example, junior team members may handle more straightforward tasks like creating data extracts and conducting simpler analyses, while more senior team members conduct more complex analyses, assist in troubleshooting, or perform quality control. The fee structure for physical extracts makes this type of division of labor feasible, whereas the VRDC's large per-seat fees often make it necessary to assign all tasks to a single team member.
- Preventing efficient allocation of effort over time and use of CMS data for longer projects. Many projects benefit from being conducted over multiple years. Researchers can face competing demands on their time that require them to cease work on a project and return to it later or encounter obstacles that require them to pause data analysis while finding a solution. The publication process can also require data access over an extended period. In some fields, especially economics, papers can undergo multiple rounds of review over a period of years before being accepted for publication, and researchers frequently need data access to produce updated results at several stages during that process.³ Because researchers using physical extracts do not face additional fees for longer projects, they can adopt whatever timeline makes sense for a specific project. By contrast, the VRDC's per-quarter fee structure forces researchers to compress project timelines and may make some types of projects that need to be conducted over a long period infeasible.

³ The VRDC does allow a single 30-day data access window after the conclusion of a project that is intended to assist researchers in responding to requests that arise during the peer review process.

While the physical extract fee structure does have important advantages for some users, there is no question that the VRDC fee structure is better for others. For institutions that conduct a low total volume of research or lack a large existing library of CMS data, the cost of acquiring physical data is often prohibitive, and the VRDC is a better (though still relatively expensive) option. Indeed, we rely on the VRDC in our own work for precisely these reasons. Nevertheless, for large institutions with large existing data libraries (which likely account for a large fraction of all research using CMS data), the current physical data fee structure is highly functional, and shifting to the VRDC fee structure would likely substantially reduce research output.

Recommendation #1: Offer institutional- or, ideally, funder-level “site licenses”

While there is little public data on CMS’ cost structure for operating the VRDC, we suspect that the fees that CMS charges for VRDC projects and seats greatly exceed CMS’ *marginal cost* of allowing additional access (because many costs of operating the VRDC, including creating extracts, designing computing infrastructure, and establishing policies, do not scale with usage). If this is true, then it is likely feasible to reform the VRDC fee structure in ways that would at least partially replicate the advantages of the fee structure currently used for physical data access.

Concretely, we envision that CMS would allow institutional users to purchase a VRDC “site license” from CMS that would allow them to pay project and seat fees close to CMS’ marginal cost of providing the additional access.⁴ The annual price of a site license might be set at a level similar to the current price of acquiring a year’s worth of physical extracts of Medicare and Medicaid data. Under this arrangement, the marginal price of data access would remain closer to where it is for physical data, but CMS could still collect the revenue needed to cover its costs.⁵

While offering institution-level site licenses would be an important improvement over the current VRDC fee structure, it would be even better to create similar arrangements above the institutional level. A large fraction of research using restricted-access CMS data is funded by the National Institutes of Health, the National Science Foundation, and a small number of private foundations. CMS could consider negotiating arrangements with one or more of these funders under which research supported by these entities (or even all research projects) would benefit from discounted fees in exchange for a periodic lump-sum payment from the funder. Unlike the institution-focused approach, this approach could also improve data access for researchers at less-resourced institutions. Additionally, since more of these funders’ support for data access would flow directly to CMS rather than indirectly through grants to researchers, the funders would likely pay less toward overhead costs at researchers’ institutions.

⁴ Under this approach, researchers would still face a cost of starting new projects, adding new seats, or operating seats or projects for longer periods, so they would still have clear financial incentives to limit use of CMS data to cases where there is an actual need for those data and to discontinue access promptly when that need ends.

⁵ While this arrangement would only directly benefit researchers with the resources to purchase a site license, researchers at other institutions might benefit as well since this option might allow CMS to collect more aggregate revenue from well-resourced institutions, a portion of which could be used to reduce base VRDC fees.

Recommendation #2: Ensure that CMS' data access processes operate efficiently

While improving the *structure* of the fees charged in the VRDC can help facilitate data access, the overall *level* of fees matters too. Because CMS' fees are set to recover its costs of providing data access, keeping fees at an appropriate level requires CMS to operate its data access processes efficiently. This includes processes for: administering data access; creating extracts and associated documentation; and designing, building, and maintaining the VRDC's computing infrastructure. Without appropriate cost discipline in CMS' operations, it will likely be difficult to fully (or even largely) mitigate the downsides of shifting more research into the VRDC.

Because we lack detailed data on CMS' costs to provide data access, it is difficult to directly assess how much room there is for CMS to become more efficient. However, the fees CMS charges for VRDC access appear large relative to the cost of the needed computing resources and the complexity of the relevant administrative processes. Consistent with the discussion above, we suspect that this partly reflects the fact that the fixed costs of operating the VRDC are being spread across a relatively narrow user base, in which case these issues may be mitigated by increasing the number of VRDC users or through the "site license" approach we proposed above. However, we suspect that there are also important opportunities for CMS to become more efficient.

As just one narrow example, consider how ResDAC processes information on what data files researchers are seeking to access. During the application process, researchers must list this information in two different places, which can be time-consuming and error-prone given the number of different files CMS makes available. Once submitted for review to ResDAC, those lists are then (apparently) manually checked against one another and, assuming they match, fed into an additional manual process to generate an invoice. Considerable effort could be saved on both the researcher and ResDAC side by collecting information once in a structured format (e.g., via a web form) and then using that structured information to automatically generate an appropriate invoice. Adopting more automated processes here and in other aspects of the ResDAC application process could also facilitate researcher "self-service" that is not feasible today.

Recommendation #3: Reconsider whether CMS must recover its full costs through fees

CMS currently sets fees that recover its full costs of providing data access. But we believe that CMS should reconsider this stance. Research using CMS data has substantial scope to improve the functioning of CMS programs and the health care system overall. As such, subsidizing these costs may be appropriate, although we recognize that the resulting benefits would need to be weighed against other uses of CMS' administrative funds. (We also recognize that it would likely require CMS to seek an exception from the federal rules that generally govern user fees.)

Were CMS to implement this type of subsidy, there would be a particularly strong rationale for subsidizing the fixed costs associated with providing data access (and charging researchers only the marginal costs). Setting fees above the marginal cost of data access is highly likely to deter some socially valuable research projects. This is the case even if researchers and their funders place appropriate weight on the social benefits of research when allocating research effort since they likely deprioritize research that is relatively more costly to perform.

Recommendation #4: Create lower-fee options for low-usage seats/projects

In addition to the changes outlined above, we believe that CMS should allow seats or projects with VRDC usage below a specified level to pay lower fees. We believe that these options would help address two problems with the existing VRDC fee structure. First, a lower seat fee for low-usage seats would make it feasible for team members who are not the primary programmers on a project but who do need occasional access to raw data (e.g., to assist in troubleshooting or quality control) to do so. Second, setting a lower project fee for periods when a project experiences low usage would allow researchers to spread project work over time in a more sensible way and, in particular, better accommodate research projects that go through periods of inactivity.

A potential concern with this approach is that some researchers who currently pay for full access would shift to these lower-fee options. If that shifting was large relative to total take-up of the new options, then CMS might need to raise fees for its existing access options, which could partially or fully offset the benefits of creating these new options. But we believe that CMS could likely largely avoid this problem by setting sufficiently low usage limits for the new access options.

We also note that the view-only seat option that CMS broached in its request for information would be an imperfect substitute for the low-usage seat option we envision here. Many of the use cases targeted by our proposal require at least a limited ability to manipulate the underlying data.

Thank you for the opportunity to comment on these issues. We hope that this information is helpful to you. If we can provide any additional information, we would be happy to do so.

Sincerely,

Loren Adler
Associate Director & Fellow, Center on Health Policy
Economic Studies Program
The Brookings Institution

Matthew Fiedler
Joseph A. Pechman Senior Fellow in Economic Studies
Center on Health Policy
Economic Studies Program
The Brookings Institution

Richard G. Frank
Leonard D. Schaeffer Chair in Economic Studies
Director, Center on Health Policy
Economic Studies Program
The Brookings Institution