From:	Samberg, Matthew
To:	EBSA MHPAEA Request for Comments
Subject:	UPMC Health Plan Response to Technical Release 2023-01P
Date:	Tuesday, October 17, 2023 4:07:26 PM
Attachments:	UPMC Health Plan - Comments on MHPAEA Technical Release.pdf

Please find attached the comments of UPMC Health Plan and the UPMC Insurance Services Division to Technical Release 2023-01P.

Thank you,

Matt Samberg
Sr. Dir. Behavioral Health Regulatory Affairs & Associate Counsel
UPMC Health Plan
600 Grant Street, 55th Floor
Pittsburgh, PA 15219

UPMC
LIFE CHANGING MEDICINE

UPMC HEALTH PLAN

U.S. Steel Tower
600 Grant Street, 55th Floor
Pittsburgh, PA 15219
T 412 454-1329

October 17, 2023

Xavier Becerra
Secretary
Department of Health and Human Services

Lisa M. Gomez
Assistant Secretary
Employee Benefits Security Administration
Department of Labor

Douglas W. O'Donnell
Deputy Commissioner for Services and Enforcement
Internal Revenue Service

Re: Comments on Technical Release 2023-01P

Dear Secretary Becerra, Assistant Secretary Gomez, and Deputy Commissioner O'Donnell:

UPMC Health Plan, Inc. and the integrated companies of the UPMC Insurance Services Division are pleased to submit the following comments in response to Technical Release 2023-01P, Request for Comment on Proposed Relevant Data Requirements for Nonquantitative Treatment Limitations (NQTLs) Related to Network Composition (the "Technical Release"). These comments are being submitted in conjunction with our broader comments on the Proposed Rule regarding Requirements Related to the Mental Health Parity and Addiction Equity Act (RIN 1545-BQ29, 0938-AU93, 1210-AC11) ("the Proposed Rule"). Please see UPMC Health Plan's comment letter for more background on UPMC Health Plan, and for our broader comments about the Proposed Rule. Our comments in response to the Technical Release are organized below in the order that the relevant issues appear in the Technical Release.

Comments in Response to Section IV (Relevant Data To Be Collected and Evaluated Related To Network Composition)

A. Out-of-Network Utilization

While data on out-of-network (OON) utilization can be a useful guide for health plans in evaluating certain aspects of their network composition, we are concerned that simply comparing OON utilization between mental health and substance use disorder (MH/SUD)

services and medical/surgical (M/S) services is an overly simplistic use of that metric, which could lead to unhelpful and misleading results. Below, we propose some possible alternative measurements.

Comment 1: Analysis of total utilization of MH/SUD services across plan types (e.g., PPO vs. EPO) would provide useful data.

The Departments several times cite a 2019 Milliman report, which showed that in commercial health plans, OON utilization of MH/SUD services was multiple times that of M/S services.[1] While we find the methodology in the Milliman report to be objectively sound, we disagree with the conclusion that this result actually reflects limitations of health plan networks.

The Milliman report only analyzed PPO plans – plans where the enrollee has the ability to self-direct their care out-of-network (albeit usually with higher cost-sharing). While this focus on PPO plans was a necessary methodological limitation, it also makes it difficult to draw conclusions from the results. In our view, the obvious next question from the Milliman report is: what about people enrolled in plans that do *not* allow OON self-direction (i.e., HMO and EPO plans)? Are individuals enrolled in those plans forgoing MH/SUD services due to a lack of in-network providers? And with respect to PPO enrollees, what is the effect of self-selection bias (i.e., are PPO enrollees selecting PPO plans because they already intend to use OON providers or place higher value on the availability of OON providers)?

Based on our experience, our conclusion is that members in EPO and HMO plans are getting similar amounts of MH/SUD services; they are just getting those services *in-network*. It therefore should not be assumed that people in PPO plans are self-directing out of network because there are no in-network (INN) options available to them. The decision to seek services out of network could be driven by a number of factors, including avoidance of the stigma that still surrounds MH/SUD treatment, the amenities advertised by OON treatment facilities, and likely other factors. Determining the reasons that such individuals self-direct out of network is an inquiry worthy of further investigation.

Rather than relying on OON utilization as a measure of network adequacy, we propose that a more useful comparative analysis would be to look at total utilization of MH/SUD services by members in PPO plans versus total utilization of MH/SUD services by members in plans that share the same networks but which do not have OON benefits. If members in PPO plans and EPO plans with the same networks utilize the same amount of MH/SUD services, then it may mean that the network itself has adequate capacity, regardless of the OON utilization of

[1] https://www.milliman.com/-/media/milliman/importedfiles/ektron/addictionandmentalhealthvsphysicalhealthwideningdisparitiesinnetworkuseandproviderreimbursement.ashx

members in PPO plans. While this measurement would not be a perfect proxy for network capacity, it would nonetheless provide useful and actionable data.

Comment 2: Analysis of OON utilizations should focus on local, in-state treatment.

If out-of-network utilization is nonetheless used as a basis of comparison, we recommend that the relevant comparison focus on out-of-network providers within a plan's service area and *not* out-of-network providers outside of a plan's service area.

For example, UPMC Health Plan sells insurance coverage only to individuals or groups located in Pennsylvania. We therefore view OON utilization very differently depending on whether it is within Pennsylvania or outside of Pennsylvania. Given the burdens of traveling for medical care, if a member receives services outside of Pennsylvania, it is likely for reasons that do not reflect on our network: the member may be traveling or otherwise temporarily outside of our service area; the member may be seeking care from a specific subspecialist who is an expert on their condition; the member may be seeking "destination" treatment from a specific facility.

High OON utilization of non-participating providers *within* Pennsylvania may be more likely to be indicative of network gaps in a particular geographic region or in a particular specialty, and that is something our Network department will take into account while looking for network expansion opportunities.

Based on our experience, comparing levels of in-state OON utilization between M/S and MH/SUD would be a more accurate reflection of a plan's network composition practices. We recognize that "in-state" versus "out-of-state" is not a perfect proxy for whether a member is traveling for care – and other, more flexible measures such as regions or statistical areas may be more appropriate for a given plan – but such metrics would be both feasible to measure and informative.

B. Percentage of In-Network Providers Actively Submitting Claims

Comment 3: Measuring the percentage of INN providers actively submitting claims is not an adequate stand-alone metric.

Measurement of the percentage of INN providers actively submitting claims ("active providers") does not provide an adequate standalone metric by which M/S and MH/SUD networks can be compared.

First, the word "actively" obscures a wide range of behavior. Even if there were a defined threshold for what it means to be "actively" submitting claims, measuring providers against this threshold would not serve to measure a plan's network composition – both because of the artificiality of any proposed definition of "active" and also because of the lack of accounting for the number of members being seen by each provider. A plan in which 80% of MH/SUD

providers regularly submit claims but see an average of 10 plan enrollees per week has a much more robust network than a plan in which 100% of MH/SUD providers each see one plan enrollee per week.

Second, the percentage of providers actively submitting claims is only a useful measure when coupled with an indicator of the network's size. Consider a plan that determines that 90% of M/S providers are actively submitting claims, but only 80% of MH/SUD providers are active providers. If the plan has significantly more MH/SUD providers than it has M/S providers, then the plan may actually have *greater* access to MH/SUD providers; the numerator alone is insufficient without the denominator.

Using "percentage of active providers" as a metric appears based on an assumption that health plans build their provider networks to some fixed target of providers and then cease all further efforts to recruit providers into the network. However, this is not how network composition works in practice. UPMC Health Plan, for example, has no such target; it is constantly expanding both its M/S and MH/SUD networks. In a competitive health insurance landscape, continually expanding the pool of available providers is an important business strategy.

In order to make the percentage of "active" INN providers a useful metric, we recommend that it be utilized only as one part of a broader analysis that looks at whether a plan has a sufficiently robust provider network to meet the needs of its members. However, we also note that this sort of broader analysis would be an incredibly complicated task. We discuss the details of such an analysis in Part C, below.

Comment 4: Measuring the percentage of providers accepting new patients has the same limitations as measuring the percentage of active providers.

The Departments ask whether it would be helpful to measure the percentage of providers accepting new patients, either in place of or in addition to measuring the percentage of active providers. Measuring the percentage of providers who are accepting new patients is a complicated task, not only because compiling accurate data is difficult, but also because the question of whether a given provider is accepting new patients can be both a non-binary inquiry and also a moving target based on a variety of factors. It is possible to look at proxy measurements – such as the percentage of providers who have recently billed a claim with an "initial visit" CPT code – but proxy measurements are not an actual measurement of the status of providers. We encourage the Departments to consider the various issues raised and comments received on the challenges of collecting and validating this data element as part of the Fall 2022 CMS RFI on a National Directory of Healthcare Providers & Services (NDH).

In addition, the percentage of providers accepting new patients has the same limitations discussed regarding the percentage of active providers: it does not account for the number of providers in the plan's MH/SUD network, and it does not account for the number of patients

being treated. Both metrics are methods of *adjusting* a measure of a plan's network composition, but neither provides a method of actually *measuring* that network composition.

C. Time and Distance Standards

Comment 5: Time and distance standards have been a useful baseline evaluation, but telehealth is rapidly changing the status quo.

Time and distance standards are a good measure of baseline compliance with respect to provider network composition, but it is important to note the limitations of these metrics. Specifically, time/distance standards do not take into account either the supply of or demand for providers in a given location.

In addition, in rural areas, it is sometimes the case that meeting baseline time/distance standards is not feasible simply because there are not enough providers in a geographic area. Indeed, state regulators recognize this fact, and often allow waivers for failure to meet network adequacy requirements in rural areas. This would also have to be accounted for in any parity analysis; there may be rural areas where there are just enough medical/surgical providers to meet basic access standards, but not enough MH/SUD providers to meet those same standards.

Finally, it is an open question whether time and distance standards are even a useful benchmark anymore, as so many modalities of care have incorporated utilization of telehealth. This is particularly true for psychotherapy, which does not have the physical location or equipment requirements of many medical/surgical practices.

Comment 6: Provider-to-Enrollee ratios would be a useful, objective, and independent standard, but there is no generally accepted source for such ratios, and there are many complications.

The Departments ask: "Should the Departments require plans and issuers to collect and evaluate the ratio of providers to participants, beneficiaries, and enrollees (also known as provider-to-enrollee ratios)? Are there models, either from Federal network adequacy or state network adequacy requirements, that could inform such a measure?"

This is a difficult question for numerous reasons:

- There is no generally accepted source for appropriate provider-to-enrollee ratio targets. Moreover, these targets would be highly sensitive to the specialty (or subspecialty) in question. Therefore, separate measures would have to be developed for MH/SUD specialties and M/S specialties so that plans' success in hitting these targets can be adequately compared.
- Provider-to-enrollee targets would be rough heuristics at best. For professional providers, the number of providers would not truly show the network capacity unless adjusted based on how frequently each provider is billing for services. For facility

providers, the number of providers does not provide useful data unless one also knows the total number of beds in the facility.

- In determining provider-to-enrollee ratios, it is important to determine the appropriate breakdown of providers by license type and level of training. MH/SUD providers are not a monolith: it is important to have a range of provider types, including psychiatrists, psychologists, and other non-physician providers such as social workers or licensed counselors and therapists in a network. There are important decision points regarding which of these groups should be aggregated or disaggregated for the purpose of determining a target provider-to-enrollee ratio.
- Health plans primarily determine whether there are sufficient providers to meet member needs by looking at historical utilization data. However, as the Departments have noted, if one starts with the baseline assumption that MH/SUD services have historically been underutilized compared to the need for such services, relying on historical data may undercount the need for MH/SUD providers.[2] At the same time, it is not clear whether or how standards for measurement could actually be made objectively "better" or more reliable if developed through a process that disregards historical utilization data.
- More than any other area of services, utilization of MH/SUD services is in significant flux, so any provider-to-enrollee ratio targets are likely to need regular updating.
- Even if the Departments can develop appropriate provider-to-enrollee ratios, plans still require direction as to the appropriate comparisons between M/S provider types and MH/SUD provider types. Because the goal is to assess parity within a given network, failing to meet a prescribed provider-to-enrollee ratio for (e.g.) psychotherapists in a geographic area would *not* be evidence of non-compliance if the ratio for the comparable M/S provider type (e.g., physical/occupational/speech therapists) was *also* not met.

In order to make provider-to-enrollee ratios a useful measurement, we recommend that the Departments:

(1) Develop uniform target provider-to-enrollee ratios for various M/S and MH/SUD provider types for the purpose of measuring MHPAEA compliance, based on independent, methodologically validated sources and taking into account all of the issues described above.
(2) Create consistent, universal rules for the appropriate M/S provider type comparators for each MH/SUD provider type.
(3) Require plans to compare results for M/S and MH/SUD providers based on the established comparator groups.

[2] *See* 88 F.R. 51573 ("[T]he proposed rules would prohibit plans and issuers from relying on historical plan data or other historical information from a time when the plan or coverage was not subject to MHPAEA or was in violation of MHPAEA's requirements….").

(4) Require plans to evaluate their performance relative to the new ratios, adjusting appropriately for "inactive" providers or providers not taking new patients.

Comment 7: Even if independent provider-to-enrollee ratios can be determined, they can only operate as a safe harbor and should not be used as determinative of non-compliance.

We recommend that if provider-to-enrollee ratios are used as a benchmark for compliance, they *only* be used as a safe harbor; they should not be used as determinative of non-compliance. Especially in rural areas, there are always going to be counties where meeting target ratios is impossible because of a dearth of relevant providers in the region. We recommend that in any case where a finding of MHPAEA noncompliance is based on failure to meet a provider-to-enrollee ratio benchmark, the Departments have a process for the plan to justify or explain its failure to meet the benchmark.

Comment 8: Compiling data on wait times is not feasible.

The Departments ask: "Are there other measures, such as wait times, that should be used to determine whether NQTLs related to network composition are designed and applied in compliance with MHPAEA?"

Wait times for seeing a network provider would be a useful measure of certain aspects of network composition. In particular, it would allow a fairly direct comparison to M/S providers (where it is not uncommon to wait months to see particular specialists). However, our concern is that measuring wait times is not practical or feasible, and it would be a constantly moving target. The only way to determine wait times would be to rely on: (1) self-reporting from providers, (2) anecdotal data from members, (3) inquiries from health plan network departments, or (4) "secret shopper" calls, where a health plan representative attempts to schedule an appointment without revealing that they work for the health plan. While together these methods can provide useful information about provider availability, they do not constitute a scalable or accurate system. This data also suffers from some of the challenges associated with data on practices accepting new patients: it is not a singular inquiry, as different patient needs may result in different wait times for an appointment, and the human element of healthcare scheduling means that this is also a constantly moving target. Moreover, even if plans were able to compile this data, it would quickly become out of date.

D. Reimbursement Rates

Comment 9: Provider reimbursement is not an outcome measurement

As UPMC Health Plan discusses in its broader comment on the Proposed Rule, with respect to network composition, reimbursement is not an *outcome* measurement at all. Reimbursement is not a measurable effect of a plan's network composition. If anything, it is precisely the opposite: a plan's network composition would be a measurable effect of a plan's reimbursement.

Regulating one of the inputs of network composition as an "outcome" is a category mistake, which will lead to significant confusion among plans and regulators.

Comment 10: When reviewing CPT codes, the Departments should also include physical/occupational therapy codes, as well as E&M codes for initial visits.

If the Departments do wish to measure reimbursement rates as part of their analysis, we have several concerns about the codes that the Departments identify for comparison. The Departments note that analyzing provider reimbursement may involve reviewing specific CPT codes and "comparing them to each other, as well as to Medicare rates…, or a similar benchmark." For example, the Department lists CPT codes 99213 and 99214 (Evaluation & Management or "E&M" codes) and 90834 and 90837 (psychotherapy) as relevant codes for evaluating plan compliance.

As an initial note, CPT Codes 90834 and 90837 are psychotherapy codes, meaning they will *only* be billed by MH/SUD providers. Thus, the only way to use them as a basis of comparison would be to compare rates against a benchmark (such as Medicare rates) and to do the same for an analogous physical health service. In this case, our analysis is that an analogous service would be another rehabilitative/habilitative therapy that is most often performed by non-physicians, such as physical or occupational therapy. Therefore, codes such as 97110 (therapeutic exercises) or 97530 (therapeutic activities) may be the best analogy to psychotherapy, if both sets of codes are being compared against a benchmark.

In contrast, E&M codes are billed by both M/S and MH/SUD providers; it would therefore be possible to directly compare M/S and MH/SUD reimbursement for these codes. However, it would be important to include *initial* visits – such as 99203 and 99204 – as well as *return* visits (99213 and 99214) in any analysis. In particular, our experience is that different provider types – depending on their business model – may be willing to accept lower reimbursement for initial visits in exchange for higher reimbursement for return visits, or vice versa. Especially given this fact, a thorough comparative analysis should look at *average* claim reimbursement over the course of a year, instead of looking at a particular code in isolation. Suppose that a code-by-code analysis shows that MH/SUD providers are paid 20% less than M/S providers for initial visits, but 10% more than M/S providers for return visits. In order to evaluate parity in this circumstance, one would need to look at the average reimbursement per visit, perhaps by performing a weighted average of 99203 and 99213. This scenario may be further complicated by other idiosyncrasies of various providers or provider types that may also have a significant impact on overall provider reimbursement, such as whether the provider frequently provides additional services in conjunction with an E&M visit, and the aggregate value of such services.

Comment 11: Reimbursement rate analysis should take geography into account.

We recommend that any analysis of reimbursement rates take into account geographic differences. However, geographic differences are going to be unique to each plan based on that

plan's service area, and plans must have leeway in explaining the geographic differences that they consider when setting rates.

E. Aggregate Data Collection

Comment 12: Data on self-insured plans may not be reliable or useful unless aggregated with data from other plans.

We are concerned that it is impossible for self-insured group health plans to perform comparative analyses using their own data. Many of the measures discussed above are based on averages, and many group health plans are too small for those averages to be statistically valid, or for any differences in those averages to be statistically significant.

For example, when UPMC Health Plan's Actuarial Services Department generates estimates of expenses for self-insured groups – including the estimates used in performing the financial requirement testing required by MHPAEA – it is rare that a single group's data reaches a threshold of actuarial credibility such that it can be used for predicting future spending. More often, that data must be combined with data from similarly situated groups in order to have enough data to reach actuarial credibility.

Accordingly, we recommend that the Department allow leeway for self-insured group health plans to work with their TPAs to get an appropriate data set that can be used for comparative analyses, and recognize that the data set may not be just the group's own data.

Comments in Response to Section V (Future Potential Federal Enforcement Safe Harbor for NQTLs Related to Network Composition)

Comment 13: Analysis of Parity compliance should take into account good faith efforts to meet the applicable standards.

The Departments ask: "To what extent should plans and issuers be able to show that they have made reasonable, good faith efforts to meet the applicable standards to qualify for the potential enforcement safe harbor?" We believe that enforcement of parity laws should absolutely take into account good faith efforts to meet the applicable standards; in our view, looking at good faith efforts is more in line with Congress' original intent in passing MHPAEA.

The MHPAEA provisions regarding NQTLs – and the 2013 final rules regarding NQTLs – as well as the comparative analysis provisions contained in the Consolidated Appropriations Act, 2021 focused entirely on *process*. Both Congress and the Departments recognized that achieving parity in administration of plan benefits was not a single destination to be achieved, but a constant process of evaluating and re-evaluating processes based on new data. If the Departments start looking at outcome measures as a key component of parity compliance, it is reasonable to assume that most (if not all) plans will "fail" some significant percentage of these

outcome measures at first. We believe the important question is whether health plans are undertaking good faith efforts to address any gaps identified by the Departments. Even then, such gaps might not be closed in a single year, which means that looking at year-over-year progress (and considering all relevant facts and circumstances) is also important.

Comments in Response to Section VI (Comment Solicitation)

Comment 14: It will take more than a year to establish the required data collection systems.

The Departments have asked about the timeline for plans to collect the required data. As described above, the outcome measures proposed by the Departments still need significant clarification. Some of the Departments' requested measurements (such as network capacity) may involve development (either by the Departments or by health plans) of new benchmarks that have not previously existed. In addition, given that many plans likely do not collect the requested data at the current time, plans may have to perform significant software upgrades or engage internal or external staff in significant development work to perform the necessary queries. Any measurements that impose new reporting requirements on providers will also require communication, education, changes to office workflows, and potentially renegotiation of network provider contracts; while all of these may ultimately be achievable by plans and providers, they nonetheless take time.

Finally, once data measurement systems are in place, plans will need time to evaluate the first batches of data, review them for accuracy, determine what plan processes should be updated based on that data, and then to update plan processes accordingly.

Therefore, even if a final rule is published in the near future, we anticipate that any requested data will not be fully available until at least the 2026 plan year, and therefore that it is unlikely that the data could feasibly used to measure MHPAEA compliance until the 2027 plan year at the earliest. Such a timeline also assumes the near-term availability of all comprehensive and actionable final guidance necessary for plans to undertake requisite implementation activities.

Comment 15: Reliable data on the number of providers in relevant services areas is not available.

The Departments ask: "Do plans and issuers have access to data showing the percentage of providers in relevant service areas and categories that participate in the plan's or coverage's network of providers?" The answer to this question is unfortunately "no."

Measuring the number of providers of a given type in a given geographic region is not an easy task. For example, in Pennsylvania, the Pennsylvania Insurance Department assists plans with network adequacy evaluations by comparing plan data on network providers to independent data compiled by an analytics company. This analysis is unfortunately unreliable in many cases. Specifically, identifying available providers within a given category relies on self-identification by providers, and it is not always the case that a given provider will be categorized the same

way in two independent analyses. In addition, there are frequently mismatches in the nature of data actually being collected; for example, the health plan may credential a certain provider type at the organization level (e.g., a home health agency), while a given data analytics company may be looking at the number of credentialed *individuals* in a region, meaning that the two data points are not comparable. Finally, the data by its nature reflects the time at which it was collected and therefore is not always up to date. Unfortunately, there are no good sources of independent data in this area at the current time.

Respectfully submitted,

Matthew Samberg
Senior Director, Behavioral Health Regulatory Affairs & Associate Counsel
UPMC Health Plan