Session 4 – Webinar Summary

Session Objectives:

1. Describe outcome measures relevant to HWPs.
2. Identify valid, reliable, and generalizable tools and/or methods to collect outcome measures.
3. Prioritize outcome measures to identify core measures to include across HWPs.

General Considerations for Anthropometric and Behavioral Outcomes in HWPs
- Identify outcomes of interest to program leadership and staff.
- Balance feasibility and validity of measures selected.
- Understand what stakeholders want to see and educate them as to why certain measures are included.
- Set expectations, particularly with children's growth and development.
- Consider non-anthropometric outcomes.
- Use intermediate outcomes given many of these programs are fairly short.
- Encourage accountability of measurements; this is a culture that needs to be developed so program staff are in tune to accurate measurements to uphold the validity of the data.

Anthropometrics - Examples
- We tend to think about height and weight first.
- Waist circumference and other measures of body composition may also be included in programs.

Anthropometrics – Reliability
- Staff training is critical to ensure everyone is conducting similar procedures, particularly if there are multiple sites for the program.
- Consider the setting that the program is being held in to ensure measures are appropriate.
- If the program is being held at an offsite location, think about the portability of equipment.
- Avoid errors in measurement.
    - For example, having a protocol to measure twice can avoid errors.
- Avoid inconsistent methods of measurement by ensuring staff are trained consistently and are consistent in how they take measurements across program participants.
- Considering when to measure is important.
    - For example, height and weight baseline may be an outcome of interest but do not need to be the first thing that happens when the participant arrives.

Anthropometrics – Logistics
- Cost of equipment
- Time to conduct measurements
    - For example, consider what other participants are doing while each individual is being measured/assessed.
- Frequency of measurement
    - Often this is pre- and post-program but consider if it is feasible or desirable to have more time points during the program.
- Consider not overemphasizing weight.
    - Allows participants to focus on other aspects of the program
    - Measuring weight in kg is one protocol that can help reduce the focus on "pounds."

- Consider privacy, based on the setting.
- Consider acceptability of measurements.
  - For example, waist circumference can be seen as a less acceptable measure.
- Consider whether caregiver or sibling measurements are desirable or feasible to track.

POWER Site Guidance
- The Pediatric Obesity Weight Management and Evaluation Registry (POWER) offers guidance to sites on height and weight measures, among others.
  - For example, guidance for weight ensures there are protocols for removing shoes and extra layers of clothing, emptying pockets, and removing heavy jewelry.

Measuring Anthropometric Change Over Time
- If you have height and weight measurements, you can calculate outcomes such as raw BMI, BMI z-score (BMIz), % of the 95th percentile.
  - There are limitations to BMIz, but much of the literature includes BMIz and may be useful for comparison purposes.
  - All the measures can be calculated with accurate and timely height and weight measures during the program.
- Waist circumference can also be measured for change over time.
- Weight gain velocity is slightly different because you need the pre-program weight trajectory/velocity that can be quantified as pounds gained per month or year for example.
  - You work to slow the weight gain velocity during the course of the program.
  - Weight gain trajectory slides can be helpful when explaining weight change outcomes to stakeholders. They show four possible outcomes:
    - No trajectory change—BMI percentile, BMIz, and % of the 95th percentile increase
    - Trajectory altered—weight gain velocity slowed, BMI percentile and % of the 95th percentile maintained, but BMIz decreases
    - Raw BMI maintained—BMI percentile, BMIz, and % of the 95th percentile decrease
    - BMI reduced

Program Results
- Baseline anthropometry of means and standard deviation for raw BMI, BMI percentile, BMIz, waist circumference, and percent body fat in a table or figure can be helpful to show where patients are starting as a group.
- When presenting outcomes over the course of a program:
  - You can show a percentage of participants who reduced their BMI percentile (i.e., reduced BMI trajectory) and show a table with BMI, BMIz, BMI percentile, and percent fat at baseline and follow-up.
  - To factor out the program participants who had a more significant outcome, you can show the percent of participants who had reduced raw BMI and present changes in BMI, BMIz, and BMI percentile a table.

Clinically Meaningful Change
- This can be:
  - Percent of participants who reached a certain threshold of improvement in BMIz or % of the 95th percentile

- o U.S. Preventative Services Task Force defined this as 0.2 reduction in BMIz in their latest statement
- o POWER defines this as a 5% reduction in the % of the 95th percentile

Physiological/Metabolic - Examples
- Heart rate (resting or heart rate recovery from exercise)
- Blood pressure at rest
- Measures of lipids, glucose, and hemoglobin A1c are less feasible to collect in community program settings

Physiological/Metabolic Logistics
- Staff training
- Ensure you have the right equipment.
    - o For example, having the appropriate cuff sizes to measure blood pressure
- Requirements for lab work
    - o Need a relationship with an ordering medical provider
    - o Need to determine how to collaborate with someone who has this data and how to share it
- Consider whether you send the participant to a lab or if it's point-of-care.
- Determine whether participants are fasting or not.

Physiological/Metabolic – Effort Dependent Examples
- Measures of fitness
    - o 12-minute walk/run is a test that starts and ends at the same time, and you keep track of how far participants go; this may be more feasible in a group setting.
    - o FitnessGram Pacer is a test that is harder to do in a group setting.
    - o Heart rate recovery
- Measures of strength such as grip strength can be used.
- Measures of motor skill development
- Spirometry is not a common measure and may be less feasible.

Program Results
- Be aware that many participants will have normal ranges of metabolic measures at baseline.
- Mean or percent changes in blood pressure, lipids, glucose, etc. can show results of all program participants or changes among participants with elevated risk.

Lifestyle Behavior Change Measures
- Includes measures of nutrition, physical activity, and sleep
- Data on these measures can be gathered through self-report questionnaires, accelerometer/pedometer data, and patient logging.
- Examples of tools to assess these measures include:
    - o Family Nutrition and Physical Activity Assessment (FNPA)
        - ▪ On a scale from "almost never" to "almost always"
        - ▪ Scoring can show sub-scores for different behaviors
        - ▪ Total score can range from 20 to 80
    - o Feeding practices surveys
    - o Parenting style surveys
    - o 3-day and 7-day food logs
    - o Food frequency questionnaires

- o 24-hour diet recalls (self-reported or guided)
- o Some studies select questions from larger surveys.

Program Results

- Changes in behaviors can be observed based on changes in scores on assessment tools, self-reported changes, accelerometer/pedometer data changes

Associated Health Conditions

- Includes conditions such as asthma, sleep apnea, non-alcoholic steatohepatitis, and prediabetes
- Tracking these conditions over time can be complicated for community-based programs.

Measures Summary

- In order of difficulty
    - o Height, weight (multiple outcome variations)
    - o Lifestyle behaviors (from select questions to full surveys, which can take longer)
    - o Fitness (resting heart rate, 12-minute walk/run, heart rate recovery)
    - o Blood pressure
    - o Waist circumference
    - o Body composition
    - o Metabolic lab tests
    - o Strength
    - o Motor skill development
    - o Spirometry
    - o Objective measures of physical activity and nutrition behaviors
    - o Associated health conditions

General Considerations for Psychosocial Outcomes in HWPs

- Consider the possible impact on domains of functioning with your intervention with children or teens with overweight or obesity.
    - o For example, if your program is a group-based intervention, you may expect an outcome related to how children interact with others.
- Consider who is best suited to report on these measures:
    - o Child > 7 years – caregiver proxy, because reading levels are limited
    - o Child ≥ 7 years – child report and/or caregiver proxy, depending on domain of interest

Health-Related Quality of Life (HRQoL)

- HRQoL is the most common psychosocial measure; it typically asks youth about different domains of functioning:
    - o Physical, social, emotional, and school
- Multiple studies report that children and teens with overweight and obesity report significant impairments in functioning relative to healthy weight peers.
- In general, in youth with overweight and obesity, HRQoL decreases as BMI increases
- Data suggest that improvements in HRQoL are associated with participation in weight control interventions.
- PEDsQL is the most widely used measure of HRQoL.
    - o 23 items
    - o Includes the four domains
    - o Runs across the age continuum
    - o Very robust measure

- May be a fee, depending on the use

Obesity-specific Measures of HRQoL
- These measures specifically ask to what extent does one's weight contribute to challenges in the different areas of functioning.
- Measuring options include:
  - Sizing Me Up
    - Version for children where a child can report on him or herself
    - Available at no cost
  - Sizing Them Up
    - Version where a caregiver reports on their perception of the child
    - Available at no cost
  - The Impact of Weight on Quality of Life-Kids (IWQOL-Kids)
    - Includes both caregiver and teen self-report versions
    - Validated; shown to be responsive to weight change

Self-esteem/concept
- Broadly defined as thoughts and feelings about oneself
- Measures tend to be single-dimensional (in general, asks how the child or teen feels about him or herself) versus multi-dimensional (drills down into particular areas of the child or teen)
- There is fairly consistent literature suggesting decreased self-concept in children with overweight and obesity compared to healthy weight.
- A reasonable amount of literature documents increases in self-esteem or its components following participation in pediatric weight program.
  - As with HRQoL, results are mixed as to whether these improvements are related to decreases in weight status

Assessing Self-Esteem
- Rosenberg Self-Esteem Scale
  - Unidimensional
  - 10-item scale that measures global self-worth
  - Free
- Dimensional Measures of Self-Esteem
  - Self-perception Profile for Children and Self-perception Profile for Adolescents
    - Assess different domains of functioning
    - Somewhat complex in terms of administration, because children have to identify which side of the measure they're going to identify with
      - Requires some monitoring and clear instructions
    - Can administer subscales of interest rather than the entire questionnaire
    - Free

Weight-related Teasing
- Commonly endured among youth with overweight or obesity
- Perceptions of Teasing Scale (POTS)
  - 11 items
  - Very transparent (e.g., people made jokes about you being heavy)
    - "People" can be adapted to specific people such as "mom and dad"

Mood

- There are not a lot of data to support assessing pre- and post-changes in mood.
- Children's Depression Inventory 2 (CDI)
    - Considered the gold-standard
    - 28 items, but has a 12-item short form
    - Requires credentialing (requires someone with mental health background to must administer)
    - Costly
- Patient-Reported Outcomes Measurement Information System (PROMIS)
    - Publicly available
    - Includes person-centered measures
- PROMIS Pediatric Item Bank v2.0
    - Broad measures in looking at physical, mental, and social health
    - Mood has a 14-item measure
    - Free

Weight and Shape Concerns/Body Image
- Weight and body shape are different from body image
- Defined as the extent to which weight influences overall feelings about oneself
- May be more appropriate for older children and teens
- McKnight Risk Factor Survey
    - Very lengthy, but can potentially use a few relevant items
- Body Dissatisfaction Scale
    - Asks teens to rate their satisfaction with nine body parts on a Likert scale
- Figural Drawings
    - Can present a child or teen with a series of images and ask which one they currently identify with and which is their ideal
        - It assesses discrepancies between these two

"Distal" Outcomes
- Social anxiety
- Family functioning
    - Many measures of family functioning, but the data on improvements are mixed
- Social support

Cautions and Considerations
- Screening (e.g., for extreme weight control weight behaviors or loss of control eating) versus observation of change resulting from intervention
    - Some community-based programs may not include screening.
    - It's important to consider the difference between screening for something versus expecting something to change as a function of your intervention.
- Implications of assessment for follow-up
    - Avoid measures that provide diagnostic information or assess specific risk (e.g., suicidal ideation) unless there is a clear plan for responding in real time.

Treatment Satisfaction
- Important metric for documenting improvements and also recognizing places for potential improvement

Cost-Analysis

- Cost-effectiveness analysis
    - Has been looked at in some community-based weight control interventions
    - Tries to answer whether an intervention provides value relative to an existing intervention such as usual care
    - Looks at the value obtained, which is the cost relative to the health outcomes
    - Looks at what it takes to deliver the intervention
        - Facilitator/coach time
        - Space
        - Cost of materials
    - Can also look at costs to participants
        - Travel time
        - Transportation time
        - Opportunity costs for lost wages
    - Calculates cost of intervention per unit change in weight status measure for child and caregiver
- Budget-impact analysis
    - Assesses the affordability of an intervention
    - Considers payer perspective
    - Focus is on cost

Summary

- Brief assessments that are available at no-cost and are likely to show improvement:
    - HRQoL
    - Key self-esteem domains, particularly around physical appearance and global self-worth
- Satisfaction with intervention to document program strengths and inform quality improvement

Session 4 Summary Call

Participants: R. Ballard, B. Belay, T. Earle, E. Jelalian, N. Klinkhamer, M. Levy, V. Osganian, B. Saelens, W. Startbucker, A. Tindall, H. Tuttle, S. Yanovski, H. Zaganjor

The call opened with a Q&A session for participants.

- E. Jelalian asked if there are resources to do a webinar on costs and the financial sustainability of programs?
  - N. Klinkhamer indicated she would like more detail on this. She likes either the webinar or a call to brainstorm and discuss.
  - Need to look into cost-effectiveness of measures versus cost-effectiveness of the intervention
    - The NCCOR workgroup decided to save conversations for cost-effectiveness of interventions until the session on sustainability and determined that an in-depth conversation on cost-effectiveness of measures is out of scope for the CLP.
- N. Klinkhamer indicated she learned a lot during this session, including potential measures to include in their program. She noted ProActive Kids has always implemented measures based on what their funders want to see, which is typically a decrease in BMI. She wondered if, at the end of the learning sessions, the CLP could define what the industry wants to see—i.e., a guide to follow with constants that all the programs are using or can use.
  - H. Zaganjor indicated that is the goal of the toolkit the CLP aims to develop. The idea is to incorporate the key measures, as determined by the group, that can be easily collected across community-based programs. It will include a tiered approach, so programs of different capacities can go beyond the "need to have" measures.
  - B. Belay added that the CLP envisions having a range of measures, from those that are relatively easy to collect, to the ones that may be considered harder to collect.
- N. Klinkhamer mentioned that her program dropped outcomes because it was not something funders cared about. In cases like this, the CLP wondered if having some talking points or rationale for each measure would be helpful to explain why certain measures are selected over others. For example, collecting heart rate recovery can show change in a short period of time, the typical length for many community-based healthy weight programs. N. Klinkhamer indicated this would be helpful, since funders are not always clear on what the outcomess can be.
  - E. Jelalian agreed and indicated that organizations such as the American Academy of Pediatrics (AAP) are trying to broaden the conversation about moving the dial related to pediatric obesity—to think about broader outcomes; she felt this is aligned with that effort. She added having a strong pitch and empirical reasons for a measure is a great idea.
  - N. Klinkhamer noted that it is important to get feedback from programs that have used/tried to use these measures. She noted some measures sound great in theory, but may not prove practical for community-based programs to incorporate in their evaluations. She'd love to learn about cost-effective measures that can be used onsite.
- Pivoting back to the list of measures described earlier, T. Earle indicated developing the list to include the outcomes the CLP thinks are important to evaluate, each with different options for how measures can be assessed, pending the capacity of the program. This allows programs to have a choice. For example, a step-test is one of many options that can be used to assess

physical fitness. This can also help explain to funders why one option was selected over another for a particular measure.

- o B. Belay clarified that the idea is for CLP participants to recommend the easiest and most practical measures. Other measures can be provided as a reference for programs that have capacity beyond the "need to have" measures. Once the toolkit is compiled, we can identify partners to test the toolkit and provide recommendations for improvement.
    - § H. Zaganjor clarified that the toolkit will include a tiered approach for measures to include, and how each measure can be assessed; but the aim is to include those measures that participants in the group have experience with.
    - § R. Ballard also clarified that once the set of measures is decided on, it would helpful to have a group of programs in community review those measures, attempt to use them, and provide feedback.
- o B. Saelens indicated he loves having comparability in outcome measures, but worries that a tiered approach will continue to perpetuate the use of different measures, which will not allow for comparison.
    - § S. Yanovski noted that certain measures may be more universally applicable. Measures that are lower-cost and easier to do in large numbers could be designated as "common measures" to recommend. She added that you would still want other measures, recognizing that not all programs have the same capacity or needs; it's a tiered approach to the tiers.
- B. Saelens stated that his program experienced a challenge with weight status measurement being used differently across different programs. He suggested if this group can come up with a specific recommendation for reporting of weight status measures, that would be great.
    - o E. Jelalian agreed that this was a great point, and that this is an area where programs are going to be collecting the same information (height, weight, gender, age) which can translate into several different metrics. More consensus is needed on the best way to report this. She indicated that the field is currently all over the place, and this consistency would help advance the science.
    - o S. Yanovski noted the NCCOR body composition workgroup is developing a guide which will cover adiposity measurement. If height and weight are reported, whether by BMIz or percent above the 95th percentile, the raw data collection will allow for comparability across programs.
- N. Klinkhamer asked if there is a way to provide benchmarks for change, i.e., what kind of changes across each measure would be deemed successful? E. Jelalian indicated there is some consensus with some of the metrics on what is considered a clinically meaningful change. That may not be within scope of this project, but it can be thought about for some measures. S. Yanovski noted this may be more appropriate for professional societies like AAP.
- M. Levy added that, in the work he is doing, there are very few true outcome measures that are easy to track. By capturing this data, he indicated they're not capturing BMI; rather it's a BMI proxy using coding methods. They can't even see small changes in BMI, instead they have to see large changes using claims data, which takes approximately a generation to see.

Discussion of Survey Results

Responses to the Session 4 survey and responses from the in-person meeting survey show that CLP participants collect outcomes related to weight as a primary outcome. H. Zaganjor asked participants what are some of the other outcome measures collected?

- T. Earle indicated that MEND has a lengthy set of psychosocial questionnaires and food behavior patterns, along with others. At a minimum, what they collect follows CDC recommendations and guidelines for what is considered benchmarking. Their greatest kickback, which becoming much more acute now since they are implementing programs in a clinical setting, is the length of the questionnaires. Also, in many of the communities where MEND works, families generally need guidance on completing the questionnaires. Although MEND has not eliminated the questionnaires, they have recently tried to streamline some of the questionnaires. They still have full questionnaires available if sites want to collect the data, but with their newer program, Healthy Together, programs can decide not to collect the full dataset if they do not have the capability to do it. MEND is currently exploring other ways to collect some of those outcomes. T. Earle noted that it harps back to the diversity of the people commissioning programs and what they expect to see. In many cases, the psychosocial measures become as important as the BMI measures; at other times, program commissioners are not interested in these psychosocial measures. She added that she likes the tiers-within-the-tiers approach, because it allows for a discussion of other opportunities to consider, which can help educate organizers and funders. She recognizes that comparability is impeding progress in this area, noting how difficult it is to compare MEND to Triple Play, an NFL program focused on primary prevention, where MEND is considered secondary or tertiary intervention. Having core set of measures would be great.

- M. Levy said they've had experience with their board and funders to ensure they understand the difference between outcome and process measures, noting there is a lot of confusion around those. He recommended being clear on how you present these measures and ensure that funders understand the difference.

- Regarding an earlier comment, T. Earle asked M. Levy about what he uses as a BMI proxy. M. Levy indicated that they currently have 140 practices, each using a different electronic medical record system. All providers trained in Strong for Life program at Children's Hospital, and they code for BMI nutrition and PA counseling using the ICD-10 codes. When they do counseling under the Strong for Life umbrella, the codes are captured using a database and put into a larger dataset from the providers. It is a very crude method.

- E. Jelalian asked if any participants from community-based programs have capacity for online web-based survey administration with the families. Could this also work with leaders?
  - T. Earle said they used web-based portions to measure process with the leaders, but they are not yet doing this with families. Doing so could provoke many HIPPA-related questions. She thinks this is an area to look into because making it easier for the families to answer these questions would be the goal. She indicated she has an upcoming call with the YMCA management team that is adapting MEND, and she will try to find out if they've tried this.
  - M. Levy indicated they used a product called Tonic which would allow you to build any survey into it; the data collected would go into a central repository that you could run reports on. You could build a link for the tool to a standard EMR system. T. Earle said something like this would eliminate the need for staff to enter data, which would reduce delivery costs.
    - Tonic is user-friendly, but it is not open source.
  - K. Rhee suggested Qualtrics, which has a very easy-to-use interface. She used it with participants, where they were able to answer questions on their cell phones.

- Regarding overcoming challenges of survey lengths, B. Startbucker mentioned a program that saved time and prioritized measures of interest by using selected questions from a lengthier validated questionnaire. He thinks this is a reasonable approach to helping overcome the burdens participants face with questionnaires. He shared an experience of using the Healthy Habits screener (discussed during his presentation) where the funder only cared about the question regarding whether a child had a TV in the bedroom.
  - E. Jelalian agreed there are times when it makes sense to take out certain questions and prioritize others to save time, particularly with discrete behaviors. In thinking about psychosocial measures, she could imagine using one or more sub-scales from a longer measure. She added the quality of life measures are more like 20 items, and it's challenging to break them up by domains.
  - B. Saelens did not like the idea of cherry-picking certain measures due to reliability issues. E. Jelalian agreed but cited an example where you think your program is going to improve kids' self-esteem, and then you only administer the global self-concept items from a measure as a less cringe-worthy option. She indicated there is research done where these items hold as scales. Although it's not an outcome measure, she suggested it's useful to have participant satisfaction; they are not validated scales, but rather domains around which people have framed such questions.
    - N. Klinkhamer indicated her programs looks at participant satisfaction, and they would love better guidelines on what those questions should be and how to ask them.
    - B. Saelens indicated that the issue is people who drop out, do not respond to these questions. The ones who finish are happy, but what about the rest? This is true across all the measures.
    - E. Jelalian asked whether another layer of this is what strategies have people identified that are effective for getting people to respond.
      - B. Belay reminded the group N. Sherwood included some of these ideas in her presentation on process measures, but that as a part of the toolkit, we can include this.

In summary, participants felt this session was very useful. They are looking forward to having more consensus on measures, but recognize if the toolkit has too many tiers or recommendations, we may continue to have limited evaluation across programs. If programs all continue to measure different things, it will be comparisons of apples to oranges. The toolkit will need to focus on a subset of core outcome measures that be common to all community-based programs no matter their capacity. Many measures discussed may be more difficult for some community-based programs to implement, and the CLP will have to determine the level of difficulty when assigning measures to different tiers. It will also include empirical rationale to ensure funders and related stakeholders have all the information they need to support collection of those measures.