National Institute of Justice

Teen Dating Violence Measurement Meeting Summary

May 5-6, 2015
Bethesda, MD

National Institute of Justice
Teen Dating Violence Measurement

Meeting Summary

May 5-6, 2015

Bethesda, Md.

NIJ

National Institute of Justice
Teen Dating Violence Measurement
Meeting Summary

May 5-6, 2015
Bethesda, Md.

Purpose of Meeting

The purpose of this meeting was to develop a strategy for improving measurement of dating violence among adolescents. The goals of the meeting were to:

1. Identify the limitations of current measures of adolescent relationship abuse (and discuss why the answer is dependent on the purpose of the measurement);

2. Highlight emerging research and innovations in measuring adolescent relationship abuse;

3. Determine ways that the Teen Dating Violence (TDV) Concept Mapping Project can be used to improve measurement; and

4. Develop a set of recommendations for what research is needed to improve the measurement of adolescent relationship abuse.

Welcome and Introductions

Carrie Mulford — Crime, Violence, and Victimization Research Division, National Institute of Justice (NIJ)

Dr. Mulford welcomed participants and asked them to briefly introduce themselves. She then summarized the events that led to this meeting. At a meeting on TDV in 2005, participants discussed uniform definitions and developing measures but decided the field needed to develop first. In 2011, NIJ hosted another meeting, Longitudinal Data on Teen Dating Violence, which focused on studies that had collected data related to dating violence (although TDV was not their primary focus) as well as longitudinal studies that had started with a younger population to see if a measure on dating violence could be added as they grew older. At the 2011 meeting, the issue of measurement repeatedly surfaced but was tabled. This meeting intends to take up that discussion with the hope of coming to some consensus on issues related to measuring TDV. Participants had been given an article to read in advance, "Measurement of intimate partner violence: A model for developing the gold standard" (Follingstad and Bush, *Psychology of Violence*, 4(4), 369-383, 2014).

Dr. Mulford acknowledged that "TDV" is not the best terminology — "adolescent relationship abuse" is better — but TDV is used because it is recognizable.

NIJ

Presentation and Reactions to NIJ/CDC Paper on ACT-Based Measurement

Presenters: Phyllis Niolon — Centers for Disease Control and Prevention
and Carrie Mulford — Crime, Violence, and Victimization Research Division, NIJ

The paper, "Taking Stock of Behavioral Measures of Adolescent Dating Violence" (Smith, Mulford et al., in press), arose out of a collaboration between NIJ and the Centers for Disease Control and Prevention (CDC) to examine the definition and measurement of TDV. It identifies 48 TDV behavioral measures used in the 130 studies included in the analysis. Eighteen measures were created by the authors for their own studies, and 26 instruments were pre-existing. Of the pre-existing instruments, however, not all were designed for an adolescent population. Many were adapted in some way, such as shortening the length, changing response categories, replacing language (e.g., to make language gender neutral or culturally or age appropriate), or adding an item, most often related to sexual violence or electronic forms of abuse. Stalking was rarely added. The most commonly used measures were the Conflict Tactics Scale-2 (CTS-2), the Safe Dates Scale, and the Conflict in Adolescent Dating Relationship Inventory (CADRI). Although the instruments assessed the same forms of dating violence, they were operationalized differently. Some of the questions and issues raised by the study include the potential impact of phrasing survey questions and whether adaptations affect the validity and reliability of the measures; gaps in measures of sexual violence (e.g., the role of pregnancy coercion, reproductive control, and stalking); and how to take into account measures of healthy relationships, although there is no consensus on how to define or measure them.

Reaction: David Wolfe — Centre for Addiction and Mental Health, Ontario, Canada

Dr. Wolfe brought up considerations along several themes:

- How necessary is it to measure rare and extreme items (low-frequency behaviors, such as threatening with a gun)? Measurement might become unreliable if a behavior does not occur very often.

- Consider reframing what adolescents are going through at an early age (e.g., what they have witnessed) in terms of abuse.

- What is the role of girls in violence prevention? Girls report more perpetration than boys — is it violence or self-defense, or perhaps an acceptable norm for communication in relationships (e.g., pushing their boyfriends)?

- Pay more attention to context. Ideally, researchers assessing TDV ask what happened and get a good qualitative idea. Without the ability to do the ideal, researchers need proxies.

- Carefully evaluate timeframe. If the CADRI is administered 6 months to 1 year apart and dating violence happens sporadically, results may be skewed. Researchers need a good framework to help subjects understand timeframe. A dichotomous response (yes/no) may be better for some items instead of asking how many times the behavior occurred.

- Consider observations of interactions to pick up on subtleties of positive teen relationship dynamics, as in the field of child abuse research. Also, measure positive teen relationships in terms of well-being; do not only evaluate how teens treat their boyfriends/girlfriends but also how they feel about themselves.

NIJ

Reaction: Luz McNaughton Reyes — University of North Carolina, Chapel Hill

Dr. Reyes offered comments centered on three areas:

- *A clear definition of the construct to be measured.* There is a subtle difference in the CDC definitions of "dating violence" and "intimate partner violence" (IPV) in terms of harm as a consequence. In general, these ACT scales aren't qualified by "harm" as a consequence of IPV, so must "harm" occur for violence to count? Other definitions of IPV include "intention to harm" or whether the act was unwanted by the victim. These aspects should be taken into account as well.
- *Context issues, particularly self-defense and playing/teasing/joking.* How do teens define self-defense? Does retaliatory violence that results in harm count or not? Likewise, "joking" may have very harmful consequences. How do teens define "dating"? What are the definitions of informal and formal dating relationships?
- *Mode of data collection.* There seem to be major effects of mode of data collection (e.g., anonymous computer compared with in-person compared with telephone). Also, there are difficulties related to reference period (i.e., people have difficulty recalling what has happened over the past year).

Discussion

Comments from meeting participants touched upon several themes:

- *Definitions.* Kids don't pay attention to instructions or the definitions provided in the instructions. They might focus on one key word (e.g., "bully"). Have you manipulated that?
  - Response from Dr. Reyes: "We have that question but have not manipulated."
  - Studies of manipulating instructions have shown that the preamble doesn't make a difference. No one reads the directions! Whatever you want people to know, it must be in the item itself.
  - We need to separate the definitions piece from the measurement piece. We can develop tools for clinicians to use consistently, but some don't translate reliably. If you load items with too much complicated nuance, people's eyes glaze over.
- *Qualitative context.* What if teens were asked to write an explanation (e.g., "Please explain the last two incidents")?
  - We have done critical incidents interviewing. But people aren't good reporters. They will give a biased viewpoint, so there is only some degree of accuracy.
  - In a longitudinal study, teens learned that if they answer "yes," they will have to write more.
- *Separating impact from the act* (i.e., what did you do and what impact did it have?). Sometimes people feel harmed by a minor act, or sometimes they say they are not harmed, even though the act was more severe.
  - Some impacts are observable — both the victim and the perpetrator know there was an impact. Other impacts are private, and researchers can only get them from the victim. That's another measurement challenge.
- Dr. Wolfe cited two papers that are in preparation by Deinera Exner-Cortens, a postdoctoral fellow, with a list of six behavioral measures and seven attitudinal measures focusing on psychometric properties.

NIJ

Measurement Selection for Different Purposes

Three presenters talked briefly about measurement selection for various specific purposes: intervention, prevalence and screening.

Intervention: Anne DePrince — University of Denver

Dr. DePrince spoke about measurement issues pertaining to intervention, particularly where TDV was not the primary or sole focus. The measures selected depend on the focus of the intervention, she told the group. In a project on preventing revictimization among girls, she and her co-investigators combined CADRI with a more broad measure, the Traumatic Events Screening Inventory. One issue was identifying the behaviors that were the target of intervention, and considering which kinds of victimization victims can control. Because the researchers were most interested in physical and sexual violence, they focused on those scales. The researchers also grappled with matters related to severity (e.g., unwanted kissing may not count if the study is focused on more extreme victimization) and frequency. Other concerns included whether the study was capturing violence from people the girls were not in a relationship with (e.g., fathers of their children); how the girls were answering if they had a different dating partner since the previous interview; and how far to reach to assess abusive behavior. The experience of victimization was only part of the puzzle; the researchers were also interested in victims' response to violence. In summary, Dr. DePrince noted that some of the behavior-driven measures provided details that were reassuring yet did not garner enough information, whereas the qualitative data collection yielded rich information that was difficult to reduce.

Long-form prevalence survey: Bruce Taylor — National Opinion Research Center

Dr. Taylor addressed measurement issues in the context of a large national survey with TDV as a focus, the National Survey on Teen Relationships and Intimate Violence (STRiV). He and his team chose to use the CADRI for the survey because it was the best fit: It is specifically designed to be used with youth (the STRiV population are ages 10 to 18); it includes both victimization and perpetration; and it is manageable and recognizable. However, the researchers remain open to using alternative approaches and other ideas longitudinally as the study population ages. For example, researchers may include interviews with a subsample or introduce a new instrument in the upcoming fall wave of data collection so that it could be tested and compared to CADRI. Other study design decisions included: not screening out acts of self-defense/play and using a recall period of 1 year. The study asks about the current or most recent relationship only. Findings so far have shown that around two-thirds of survey respondents have experienced either victimization or perpetration.

Short-form prevalence survey: Emily Rothman — Boston University

Dr. Rothman discussed some ways she and her team have collected data on TDV when it has not been feasible to administer a long instrument (e.g., eligibility screenings). They have condensed salient items from the Safe Dates Scale into four questions by putting multiple items on the same line. They also have asked teen girls to keep a daily diary, where they are prompted to answer a few questions. Through the daily diary, they found variability between girls and within girls, demonstrating that their experience of victimization can vary. The researchers have also used timeline follow-back interview techniques, which have helped respondents walk back in time using a calendar as a memory aid. Using the calendar seems to increase event recall. Another technique used by the researchers has been a call-in study, during which participants had to call in daily to answer questions. At this point, the researchers are comparing these procedures to see how well they map and compare how many events they produced (the interactive voice response produced more events than the calendar). Finally, Dr. Rothman spoke about relationship

structure and the difficulty of identifying an accurate description. For example, questions include the word "dating," but the research population Dr. Rothman works with uses "spending time with." She noted that a lot of relationships don't even happen in person but are maintained by texting, interacting online, or spotting at school. In addition, she noted that another difficulty is multiple concurrent partners.

Discussion

Dr. Mulford summed up by noting that the presentations showed that one measure does not fit all purposes. She then invited comments from meeting participants. The exchange included comments on:

- Daily diaries
  - Daily diaries are great for pure research purposes but are limited for clinical study.
  - However, don't dismiss daily diaries just because they aren't perfect. The daily diary gives an idea of context, actions, and harm as felt at that moment, so the diary is very useful as a quick and relatively cheap tool.
  - Respondent burden could be massive if violence explodes only once every 6 months.
  - The daily diary can also act as an intervention.
- Cognitive interviewing
  - Make sure the measurement matches the intervention. Some abusive behavior in middle school is dating violence, but a lot of it is harassment. Ask youth what terms such as "sex partner" means to them.
  - An example is the question, "Has a partner forced you to have sex?" When asked what that means, respondents said "pinning me down on the bed" or "holding a gun to my head," but when asked further whether it should include acts they didn't want to do, they agreed.
  - Respondents don't want to count oral sex. Also, teens don't talk about "dating" or "girlfriend" and "boyfriend," but they say they know what adults mean when they use those terms and tell us to keep using them because they don't want adults using their language.
- Other topics
  - The idea of threshold should be considered: What is the threshold beyond which we want to count behavior? What is normative and will disappear with development out of immature interaction, and what really reflects problematic dysfunctional relationship behaviors that persist in adulthood?
  - Other methods are valuable because memory is highly fallible and relationships are very fluid and dynamic.
  - The huge difference in prevalence rates has a lot to do with sampling methods.
  - If you are doing prevention or intervention work (i.e., trying to reduce TDV as opposed to trying to understand the nuances of TDV), be pragmatic and choose something reliable and concrete so that a reporter can tell what you've done.
  - A limitation of these measures is that they are relationship centered.
  - Don't keep using less effective methodologies, even though our convention in psychology is to continue their use regardless of whether they are wrong.

NIJ

Add-Ons to the ACT-Based Scales

Six presenters spoke briefly about specific add-ons to the ACT-based scales: reproductive coercion, sexual harassment, technological forms of abuse, sexting behaviors, financial dependence/strain, and measuring healthy relationship concepts.

Reproductive coercion: Elizabeth Miller — Children's Hospital of Pittsburgh, University of Pittsburgh Medical Center

After conducting 60 in-depth interviews of young women in relationships — in which more than 25 percent talked about explicit attempts of their partner to impregnate them — Dr. Miller and her co-investigators added a 10-item reproductive coercion measure. This was later shortened to five items and will be released soon. Dr. Miller said she is often asked whether reproductive coercion is part of partner violence and answers that it is and deserves attention, but it is a distinct construct in her work. Measurement of this construct can be useful for clinic health care providers or targeted intervention (e.g., offers of emergency contraception, IUD, and so on).

Discussion

Several attendees commented that they had also included crude measures of reproductive coercion in their own studies but were grateful to have Dr. Miller's measure. A point of clarification was that "condom coercion" means *not* to use condoms, but another attendee pointed out that condom use can be complicated by the type of relationship. Some men with concurrent partners will use condoms with all girls except "my main girl" or with "my baby's mama."

Sexual harassment: Dorothy Espelage — University of Illinois

In Dr. Espelage's work, she has been modifying the American Association of University Women sexual harassment scale in her studies of middle and high school students. She reported that the scale is becoming more brief by eliminating redundancy, but gender harassment (e.g., homophobic bantering) has been added. Although a high incidence of sexual harassment (i.e., unwanted sexual commentary) is not seen, the researchers still want to measure it at its early development stage. Dr. Espelage also noted that a predominance of dismissive attitudes can create a climate for the development of sexual harassment, so she and her team now also ask questions regarding school climate. They also ask about witnessing harassment and responses to harassment, and they assess peer networks. They want to understand the precursors to sexual harassment, not simply focus on outcomes (preventing sexual harassment).

Discussion

Questions put forth in discussion brought out several other aspects of Dr. Espelage's work: The researchers did not ask about stalking but did include online sexual harassment (using items from Michele Ybarra's work). The researchers also asked whether the target was a peer or a person of special interest. Other comments included:

- It is important to measure sexual harassment in middle school, even starting in grades 4 and 5. By high school, sexual harassment victimization is a strong predictor of dating violence victimization and other bad outcomes.

NIJ

- How appropriate are these measures and constructs of peer relationships compared with dating relationships when mapped on the developmental timeline? It may be that peer and dating relationships have a lot of fluidity and are not distinct at early stages.
- If a popular kid is perpetrating sexual harassment or dating violence, others will imitate.
- Many peers will call each other names but have different motives. Some de-weaponize the word by using it themselves. Some intend to be hurtful, whereas others are joking around, not intending to be hurtful. What sounds bad in an adult relationship may not be interpreted the same way among teens — we need to be developmentally attuned.
- Name-calling might not hurt the target but, even so, there is harm to the culture. That is important to get to.
- The sexual harassment field has mostly lived in a silo. How does it connect to peer victimization?
- There is a problem with measures of flashing behavior because there is so much consensual flashing. Most don't experience it as victimization.

Technological forms of abuse: Meredith Dank — Urban Institute, Washington, DC

Dr. Dank's work used cyber dating abuse measures adapted from Picard (2007) for Teen Research Unlimited and Griezel's cyber bullying scale (2007). She and her team tweaked those measures (e.g., some words are no longer used by youth today) and separated sexual cyber abuse from other forms of cyber abuse. The researchers were only able to test high school students due to barriers in getting into middle schools. Dr. Dank noted that as technology changes, measures must change, too. She said she considers technological forms of abuse as describing another tool in the psychological coercion toolbox rather than a separate type of abuse, but it could change over time.

Discussion

Discussion covered several themes:

- How does this issue line up with dating violence in middle school years, and what does it say to kids about relationships in general? More broadly, how do these kinds of behaviors translate into more intimate relationships over time?
  - Response: "Perhaps STRiV can help answer those questions. STRiV has a younger set, and we see differences in how they shift their behavior as they mature."
- There is a parallel with smoking: Students hear how social media and technology can ruin lives, but they'll still see what they can get away with. They're willing to take a risk, no matter how much programming they receive.
- Definition issues
  - Is "pressure to send naked photos" considered psychological or sexual abuse?
  - Response: "Mapping onto the law, some would be treated as a sex offense. Some are illegal because they are sexual, and some are not because they are not meeting the law's definition of threat.
  - Coercion is a factor. Threatening to hit someone is still coercion. (It could be coded as physical or psychological.)
- What are the differences in cyber abuse toward "someone" versus toward a "dating partner"?
- In exploratory factor analysis, psychological abuse correlated well with sexual abuse. Although many people are dismissive of psychological abuse, it can have a deep impact.

NIJ

Sexting: Jeff Temple — University of Texas at Galveston, Medical Branch

Dr. Temple explained that sexting is a new phenomenon that does have some relationship with dating violence because it mimics real life. That is, having sex in real life means a higher likelihood of sexting; being sexually coercive in real life also means a higher likelihood of being sexually coercive in sexting. Negative outcomes are associated with unwanted sexting or unwanted sharing of pictures (or "sextortion"). Dr. Temple also suggested that messages to teens about the dangers of sexting should focus on the risk associated with it, commenting that the same mistake can be made in messages to teens about sexting as with telling them that "If you smoke a joint, you will die" (they know it isn't true and don't listen). Instead of a message such as, "If you sext, you will ruin your career," it may be better to compare sexting with driving without a seatbelt — it is taking a risk.

Discussion

An exchange further clarified when sexting is abusive. If a naked picture is posted but it wasn't coercive, other factors may weigh in, such as whether it was nude or seminude, images compared with text, etc. If the recipient does not want sexting, it is different than if the recipient asks for it.

Financial dependency/strain: Jennifer Copp — Bowling Green State University

Dr. Copp pointed out that measuring financial dependency/strain in young people is different than in adults. In her work, she has focused primarily on young adults in their late teens to early twenties. Because they may not have experienced the same negative life events as adults, the researchers included some subjective items (e.g., "I have potential but I'm not living up to it"). Although the relationship is often thought of as a dichotomy of positive/negative dynamics, quite a few may be mixed. When there is a mismatch in the financial relationship of a couple, abuse can happen. Greater financial dependence is related to violence — couples may fight about money or a partner who has no ambition. Instrumental support can lead to feelings of resentment and negative assessments of the future.

Discussion

Several points were brought up during the discussion that followed the presentation:

- On STRiV, financial questions are included for participants as young as age 10.
  - Financial literacy as a construct (i.e., how much kids are able to know where they stand financially).
  - One item on a current survey to help kids distinguish healthy and unhealthy relationships is "Making your partner pay is unhealthy" — we tried to get at it that way.
- Significant relationship issues are expected where there are lending issues. Financial dependence has some connection with TDV. The researchers were surprised at the high incidence of financial dependence.
- Although financial dependency is used more as a predictor, the threat of withholding financial resources is actually part of the abuse in adult relationships. This has not received a lot of attention in teen relationships, but there's definitely something to explore.
- Financial strain correlates with negative outcomes better than income or being on welfare. Some families with middle-class incomes report very high levels of stress and anxiety about financial strain.

NIJ

- Financial mismatch can go either way: Attendees noted that some women have the money or work, and their partners live off them, and girls can put pressure on boys to buy them a lot of gifts.
  - In interviews, one reason girls gave for not leaving a relationship, even if it was abusive, was because they had become financially dependent over time, getting accustomed to receiving gifts ("Maybe he yells at me or hits me, but look at all this nice stuff he does for me").
  - In a project focus group, a lot of high school boys talked about the pressure and expectations around paying the tab as well as being sexual.

Healthy relationship concepts: Sabina Low — Arizona State University

Dr. Low pointed out that healthy relationship concepts are frequently measured at some level, but the data is rarely used. Nonetheless, it is important to capture. In her work, she and her co-investigators have tended to go for depth and breadth, reflected in the mixed methods used to capture relationships. They have used interviews to explore caring behavior, commitment, and attachment and items from various sources to capture aspects such as affection, cohesion, and satisfaction. But Dr. Low emphasized that there is nothing richer than observational data to get at nuanced interactions. The researchers used positive interaction scales that included codes for both affect and content to detect micro-behaviors that highlighted the importance of repair attempts or escalating behaviors that lead to more severe, toxic behavior. Dr. Low also referenced using the Dyadic Social Skills Questionnaire put together by Dr. Deborah Capaldi.

Comment by Dr. Capaldi — Oregon Social Learning Center

I wrote that scale for the LIFT (Linking the Interests of Families and Teachers) Couples study because other scales looked too adult for these 17- and 18-year-olds.

Innovations in Measurement

Four presenters briefly addressed innovations in measurement specifically: adding a qualitative component for context, adding observational components, creating developmentally appropriate measurements, and creating experience-based sampling.

Adding a qualitative component for context: Peggy Giordano — Bowling Green State University

Before Dr. Giordano talked about the qualitative component of the Toledo Adolescent Relationship study (TARS) project she has been working on, she pointed out her amazement at how well the CTS-2 worked as a screen to find people with violent relationships (out of 102 people in wave 5 that indicated they were in a violent relationship, based on the CTS-2, 89 had a significant experience to talk about). Although she acknowledged that CTS-2 can be improved, she urged researchers not to abandon it completely. Dr. Giordano and her team have couple-level data for 50 couples. She explained that qualitative measures can step back from the violent act itself (i.e., give the actual content of conflicts). Situational context is also easier to get at with a qualitative frame. In addition, Dr. Giordano said that qualitative measures can give access to the unexpected. For example, she has seen that people do change over time and has been interested in exploring how people stop inter-partner violence. Also qualitative measures are necessary because of the nuances of gender, especially those that don't go with the dominant frame. The tricky part of qualitative measures is ending up with too much data or too much context and then figuring out how to understand it and benefit from that understanding.

NIJ

Discussion

In response to a question about how much qualitative work has been done in a survey environment online, Dr. Giordano replied that, in a new initiative, about half of the questionnaires were returned, some with good information, but they only contained a few qualitative components because respondents will not answer a lot of questions.

Adding observational components: Amy Smith Slep — New York University

Dr. Smith Slep said that she and her co-investigators have done a lot of observational work, but they have begun to apply it to teen dating relationships based on the work of Dr. Capaldi. Dr. Smith Slep described how traditionally these studies have set up a variety of ways to encourage couples to demonstrate conflict processes. Study design decisions include how long to collect data, how to set up the observation task, and how to code it to glean the information you want. In terms of how long to collect data (understanding that it is expensive), those decisions are based on the frequency of the lowest base rate of the phenomena you expect to detect. As far as setting up the tasks, one difficulty is that teen couples don't have as many longstanding issues to tap for conflict as married couples do, but Dr. Smith Slep used an interview approach to find topics, then Q-sorted them, and used the list of top unresolved issues for problem-solving sessions. She also used tasks inspired by parent-child studies, such as cooperative Lego or tangram building under time pressure, where only one partner is allowed to touch the pieces, and wraps up with a positive reminiscing task. The next set of issues is how to code the data. Dr. Smith Slep uses a coding system out of Oregon that has undergone slight revisions. Codes are based on what people are doing on a 5-point scale from high-intensity positive to high-intensity negative, but the researchers are starting to move toward interval-based coding, which allows them to look at sequences. Another option is global coding (overall impressions by the coder), which is much faster and tends to predict outcomes. Some researchers emphasize coding positives in the relationship, although this is not used much in IPV studies. In summary, although Dr. Smith Slep acknowledged that observational data is hard to get and labor intensive to collect and translate, she noted that it provides a window to contextualize behavior.

Discussion

Attendees who have done observational data collection provided further comments.

- Dr. Capaldi added that it is beneficial to have more than one kind of task because each only takes a few minutes. Also, in her work, she has codes for physical and psychological aggression and found that physical aggression was a valid predictor of later injury. She added that she has had better results when asking a couple to describe what happens when they have a bad fight rather than what happened when they fought/argued in a specific instance.
  - Dr. Smith Slep responded that in her studies, the discussion issues have to be something the couple has discussed before that are important to them (not secret topics and not areas they've given up on).
- Dr. Low added that if a subject refers to physical aggression, it is coded as physical aggression in her study. She also pointed out the difficulty of coding observational data when a couple spends the majority of their time in "neutral."

Developmentally appropriate measurement: Pamela Orpinas — University of Georgia

Dr. Orpinas described following students in sixth through twelfth grades with questions on dating and dating violence (but not sex) and examining for patterns across time (i.e., were there groups who

followed a similar trajectory over time) and whether those trajectories were related to other outcomes. For example, when two trajectories for physical aggression for perpetrators and victims were crossed, there was about 90 percent overlap: The group low in perpetration was low in victimization, whereas the group high in perpetration was high in victimization. High aggression tended to increase over time. At the end of twelfth grade, the researchers conducted interviews of teens with high aggression. Girls described a whole relationship process that went quickly from falling in love to having sex to "going downhill." The researchers also included quantitative questions to try to sequence the relationships. Although the stereotype is "started well, then went downhill," the modal response was "highs are high and lows are low," possibly indicating that there are some positives in the relationship that keep it together.

Experience-based sampling: Emily Rothman — Boston University

Previously, measures for both the physical and sexual victimization of dating violence were lumped together in one question on the YRBS, using problematic language on a survey. Drawing on the Sexual Experiences Survey and the Safe Dates Scale, Dr. Rothman and her team offered the CDC alternative questions, resulting in two questions being adopted for the 2013 Youth Risk Behavior Survey (YRBS): "How many times did someone you were going out with physically hurt you on purpose?" and "How many times did someone you were going out with force you to do sexual things you didn't want to do?" Dr. Rothman reported that the prevalence of physical dating violence remained nearly the same (10 percent) — that is, the changes in the questions did not change the percentage of physical victimization. For girls, counting both physical and sexual victimization, the prevalence is now 20 percent using a combination of the two questions. For boys, the prevalence was 10 percent (no matter if counting just physical or physical plus sexual victimization), which was an interesting change.

Discussion

In response to a question about whether the findings meant that boys reported no sexual victimization, Dr. Rothman explained that a small number of boys did experience sexual victimization, but the overlap between physical and sexual victimization did not substantially increase the overall 10-percent victimization rate. In response to another question, she clarified that the survey response options begin with "I haven't gone out with anyone in the past 12 months" to account for those who do not date.

Research on Measurement

Three presenters discussed topics related to research on measurement: how wording matters, testing for measurement invariance, and comparing current scales.

How wording matters: Sherry Hamby — Sewanee: The University of the South

Dr. Hamby declared that we know how we ask the question is important, as shown in the study "The Sexual Victimization of College Women" (Fisher, Cullen, and Turner, 2000), which demonstrated an eleven-fold difference in reported rates of rape (as compared to the NCVS) when a range of behaviorally specific screen questions was added. Although Dr. Hamby acknowledged that the first generation of measures has had great impact on understanding the problem of TDV, she urged continuing efforts to improve measures. She suggested there are unacceptably high rates of false positives and false negatives (where people are misclassified), yet measures can be easily tweaked to improve them. She encouraged researchers to experiment methodologically and push the science further; she also pointed out that when a large discrepancy in rates shows up, researchers should examine what's going on with that questionnaire. Dr. Hamby said she has a new measure that doesn't show gender parity and has multi-method convergence. She also recommended building context into the item instead of the instructions.

NIJ

Testing for measurement invariance: Liz Goncy — Virginia Commonwealth University

Dr. Goncy described her studies looking at measurement invariance (a prerequisite for comparing mean rates of teen dating violence). Critically, she treated response categories as order categorical (i.e., without assuming the same difference between categories). This accounts for severity across categories and across items, so a particular item could be weighted more, according to its severity and item properties. In her work, Dr. Goncy used samples collected from middle schools four times per year but, to reduce the burden, each individual respondent only participated twice per year, randomly assigned to two quarters. She looked across time, grade level, gender, and season. Interestingly, the data show drops in mean rates of dating aggression in January and July. It is not yet known how these drops relate to protective factors or different outcomes. She also raised the topic of construct validity (general aggression compared with dating aggression).

Discussion

In response to whether the question asking respondents to report for the past 3 months, Dr. Goncy said that, yes, they were asked whether they had a boyfriend/girlfriend within the past 3 months.

Comparing current scales: Michele Cascardi — Montclair State University

Dr. Cascardi's interest was seeing how much the CADRI and CTS-2 overlap. College students were given two instruments that had similar, although not identical, content. Dr. Cascardi restricted the CTS-2 to items that generally overlapped. Interestingly, self-reported perpetration for males was 24 percent on the CTS-2 and 15 percent on the CADRI. A similar pattern was found with females reporting victimization. Dr. Cascardi suspects that the way items are organized and the sequence of behaviors might matter. Qualifiers on the CTS-2 seem to matter more for males and increase their endorsement. Dr. Cascardi questioned whether conversations of context and thresholds matter less for males, perhaps because they are accustomed to roughhousing and certain actions seem too trivial to report. She also noted that males were more likely to respond affirmatively when items such as "slapped or pulled hair" were separated. Further questions for research include what order should questions about behavior be asked and how the meaning of behaviors varies between boys and girls. Dr. Cascardi's findings also suggest that concordance is a proxy for severity; people who were concordant reported a greater variety of actions, more frequent threats, more psychological aggression, and a higher rate of injury. However, she acknowledged that if you are looking for prevention, fluctuations don't matter as much as stopping the undesirable behavior.

Discussion

Dr. Cascardi clarified that the research was not designed to answer whether the order matters. The CADRI always came first. Fatigue was not an issue in terms of data collection.

Open Discussion

Dr. Dara Blachman-Demner and Dr. Mulford invited attendees to write down their responses to two discussion questions. Attendees then had the opportunity to share their responses with the group. Responses are grouped here by similarity of topic.

Question 1. In what ways can TDV measurement be advanced using existing data?

- Resources
  - Updating the compendium of measures that the CDC had.
  - Making the compendium electronic, and linking to new studies over time.

- Devoting special issues to TDV measurement, which could open up discussion among scholars.
- Researchers collect much more data on what is solely reported in journals and reports. Finding ways to share this data through a website or user portal to determine if there are questions already being asked that can advance measurement.
- Researchers at this meeting can share data sets with each other.

- Having a special issues/compendium combination where we can update links on the web (like a research gate).
  - Having access to data such as the National Data Archive on Child Abuse and Neglect that Cornell provides.
- The Summer Research Institute at Cornell offers a wonderful opportunity to use the data sets from this archive.
  - Finding out what data sets exist.
- Looking for data sets outside of our silo and getting access to them.
- Identifying the good data or good parts of data to do analysis on.
- If we push these data sets, there may be a concern about piecemeal publication.
- It is important to make clear how your current paper is different from the previous paper when publishing from the same data set. Journal editors require it because there isn't much added value if a paper is just tweaked or rescored.
- Comparing/combining data
  - Combining data sets that have similar measures and look at scaling work; also looking across data sets to see if there is consistency and what commonalities they have.
- Determine common items and scales used across data sets for comparison. Pool data/samples for re-analysis that would allow for comparisons. Considering problems with a low base rate, studies could have more power if data across studies is integrated.
  - Doing a mega-analysis, combining actual data into one data set.
  - Using the CADRI and combining data from other countries to come up with a measure that is more stable and up to date.
  - Crossing data sets with different samples — finding different ways to interpret measures in the samples.
  - Data harmonization: combining multiple data sets across ages, geographic locations, etc., conducting IRT approaches, and linking data back to outcomes, risk/needs, and protective factors.
- Measures
  - Somehow getting a "gold standard."
  - Performing item-level analyses to examine differential validity and sensitivity, and specificity with existing data. Ideally, we could have a "gold standard," but we could use a constellation of things, knowing which one of the existing measures best predicts the outcomes of interest.
  - Figuring out which are the best items and decreasing the burden on participants. Identify low-frequency items that can be eliminated or examined across age/gender; identify high-frequency items that should stay; identify items that could have more than one meaning; identify items most related to other constructs; and compare response categories (e.g., yes/no and 0-10).

NIJ

- Do not rely on scales that lump the various forms of victimization together (e.g., physical and sexual). If measures include terms such as "intent to hurt," "fear," or "injury," recognize that these items are not neutral or contain a specific tilt.
- Investigate the measurement invariance of specific acts (probably not enough data for anything else) as they relate to a stable higher order construct of relationship violence/abuse over the lifespan — Is a slap a slap, from a measurement perspective?
- Can we agree to use the CTS less? Guidance for how to talk about existing data that has used the CTS and other older measures, and what that means.
- Consider measures developed for college samples that could be tested and adapted for teen and middle school populations.

- Specific research angles
  - Getting a deeper understanding of co-occurrence of physical and sexual violence. (It is frustrating when studies combine the two.)
  - Looking at cultural characteristics of dating violence among teens — culturally specific dynamics and implications in certain populations.

- The literature dichotomizes a lot, and combining can be misleading — something for future studies to consider.
  - Using existing data and measures already established, examine wording and ordering issues.

- Comparing across scales for reasons other than gender symmetry
  - Examining how indicators play out differently along the developmental timeline (e.g., is a slap between 11-year-olds the same as between 22-year-olds or 45-year-olds?).
  - Exploring cultural characteristics and delineation with age; also the question of addressing the transition to adulthood.
  - Studying variations in relationships; doing Venn diagrams by age, gender, race, and ethnicity differences.
  - Using existing intervention research data to get at the mechanisms.

- Analysis
  - Resources (funding but also teaming) to do more in-depth analyses of CADRI outcomes i.e., comparing/assessing different approaches to coding and interpretation (e.g., STRiV) — latent class analyses and person-centered typologies based on CADRI item endorsement.
  - Cross-sample analyses (e.g., STRiV sample and Jeff Temple's sample) to try to match cases and then investigate and compare CADRI outcomes.
  - A consensus analysis plan across different CADRI data sets to check for consistency in patterns.
  - Use IRT or SEM to explore construct validity with measures.
  - Fund secondary data analyses and examine samples with both observational/qualitative data and quantitative measures.
  - Do psychometric studies (e.g., reliability, convergent validity, predictive validity, stability) with specific subgroups/populations (e.g., ethnic groups, LGBT youth, high-risk groups like foster care youth, adjudicated youth) versus normative groups. This will also help examine literacy and the understanding and meaning of items.

- Other topics
  - Encourage researchers to list the items and responses when they write about using a particular scale or modification of a scale.
  - Work on definitional issues and cross-walk to see how measures are similar and different.
  - Clearly define what we're measuring and for what purpose — prevalence goals are a different motivation for measurement than prevention or psychological outcomes.
  - Do more conceptual mapping.
  - Assess influences of media, climate, community, peer and social norms, etc.
  - Identify ways to capitalize on community agency data. Barriers to administering these surveys in schools and getting approval so that community-based samples or diversion programs could be approached.

Question 2. What types of new research are needed to advance our understanding of TDV measurement?

- More work on conceptual framework (e.g., define dating, dating violence, fear, wanted/unwanted, intentional, and so on) and how to incorporate it into surveys.
  - Include young people in this development: Ask them what words they use and what technology they use.
  - Validity (course, consequences) of using different typical definitions of relationship violence (i.e. act-based, act and impact-based, etc.)Need research on an agreed-upon definition of TDV.
- Experimental studies that compare different measurement approaches.
- Experimental research on wording and measurement effects.
- Instruments and sampling approaches for specific purposes (e.g., clinical screening compared with basic research).
- Convergence of self-report, partner report, daily diary/EMA, observations, interview data and implications for a) validity and b) empirically-derived screening.
  - Examining this in the context of related relationship aggression (relational, peer) to further inform measurement of TDV specifically.
- More couple-based measures so data is obtained from both partners (e.g., interactional observations).
  - Need to stop thinking about "couples" because many youth have multiple partners.
  - Need to address the question/problem of offender-victim overlap. This might be addressed through qualitative data collection to disentangle offensive and defensive behavior.
- More research on specific populations: cultural (e.g., Native American, African-American, Latino/Hispanic), child welfare samples outside traditional schools, and other forms of diversity.
  - Item interpretation among high-risk youth.
  - How does victimization and perpetration differ among at-risk and vulnerable populations?
- More research on middle school and high school students' sexual health behaviors (e.g., use of Internet porn, group sex, and so on).
- More research on fluidity of early relationships (e.g. should we only be focusing on current/most recent relationship?) This includes the context — not just what was going on directly before

episode but capture broader context and content of the disputes themselves — what are people fighting about and what kinds of conflicts most likely escalate to violence?

- Innovative ways to survey young people (e.g., texting a question every day for a period of time).
  - Innovations for non-school-based studies are needed.
  - Researchers need to share innovative ways and approaches so others can learn.
  - Use of different platforms and ways to reach youth at their level.
- Longitudinal prospective studies using mixed-methods analysis to see how the nature of dating violence may change from ages 13 to 20.
  - Also look at the validity of using different conceptualizations of dating violence and which ones reveal the most about course, consequences, psychological adjustment, future relationship issues, etc.
- More research on psychological abuse.
- Development of more qualitative approaches and comparisons across qualitative approaches.
  - Cognitive interviews focused on CADRI responses.
  - Qualitative phone interviews (Skype/in-person options) with the STRiV sample to interpret CADRI (or additional measures using future waves). Add a focus on TDV measurement to STRiV design using a subsample of baseline cohort.
- Research on the impact of violence on the relationship itself (e.g., what was the response or consequences?).
  - Studies to explore the addition of qualifiers regarding impact (e.g., physical/injury, emotional stress, harm to relationships/feelings about partner); intensity (fear; use visuals such as a picture of fists, where 1 fist equals mild, 2 fists equal moderate, and 3 fists equal severe in order to capture the level of intimidation; and consequences of TDV on the self, relationship, peer group (e.g., told to stop, made to apologize, retaliation, etc.).
- More research on the broader context (i.e., what are distinguishing features of conflicts that turn violent compared with those that don't?), similar to what is used in criminology.
- Content analysis of data collected from texting and social media. Follow-up with qualitative data collection to understand thought processes.
- Replication studies.
- Basic studies on how youth view dating (e.g., hookups).
- More studies using mixed-methods and multimethod analysis.
- How does social desirability influence differences in perpetration/victimization reporting by males and females?
- How can qualitative/context — identifying methods (e.g., interviews) — be used in a standardized way?
- Psychometrics on knowledge, attitudes and healthy relationships. Need a measure for situational context and examining "meaning-making" among youth about their abusive experiences.
- Examine the development of TDV into IPV over time into adulthood.
- How to handle the reference group (e.g., the last person you dated or the most difficult in the past year).
- Enhance memory/recall for events by determining optimal time frame.

- Explore alternatives to measuring the frequency of a behavioral occurrence, such as episode-based conflict behaviors.

- Use skip patter question logic to decrease the cognitive burden of completing the survey, and help youth "tell their story," with specific occurrences of TDV (e.g., If X behavior ever occurred, when did it happen? What else happened? How did you feel when it happened? and after it happened?

- A new approach to understanding violence and its interplay with sexual and psychological violence.

- Typologies — e.g., the number of arrestees or emergency room injuries.

- Work on contextual typologies (i.e., not the typology of the individual but of the violence itself, and the context in which it occurs) to help elucidate other issues such as severity, impact, fear, playfulness, etc.

- Need a valid, reliable, single question for college surveys and a good screening tool for health care settings with good sensitivity and specificity.

Teen Dating Relationships: Results from the Concept Mapping Project

Presenter: Dara Blachman-Demner — Crime, Violence, and Victimization Research Division, NIJ

Dr. Blachman-Demner gave a slide presentation that described a project that NIJ has been working on with colleagues from the Concept Systems Inc. and the Federal Interagency Workgroup on Teen Dating Violence. Springing from the first TDV workgroup meeting 10 years ago, the impetus for the project was to find out what youth think about dating relationships, and how this compares with adult perspectives, in the hopes of gaining improved understanding that can inform the design of TDV programming, policy, and research efforts.

The study had two parts: (1) a concept-mapping portion that used a mixed-methods approach to develop a visual map of how the group thinks about teen dating relationships, and (2) facilitated discussions with youth and adults to enrich the information and validate the map. Although the sample was not designed to be nationally representative, the researchers made it as broad as possible. Teens ages 14 to 18 and young adults ages 19 to 22 were recruited through youth-serving organizations with diverse focuses and geographic locations; adult participants were purposefully selected from among professionals in the field because they are the ones using the terms.

The focus prompt was: *"A thought, action, feeling, or behavior that teens in dating relationships might have or do is … ."* (Even though teens don't use the word "dating," they know what adults mean when they say it.) A variety of data collection methods yielded about 600 ideas to complete the focus prompt; the research team culled that list down to about 100 statements that represented the original content. Next, invited participants sorted the statements into groups according to their perceived relatedness and rated each statement on both its frequency and desirability in teen dating relationships. Cluster analysis of their input shaped a map that grouped conceptually similar ideas. The research team labeled the clusters of ideas with the input of youth. One unexpected finding was that two axes emerged: healthy compared with unhealthy relationship characteristics, and external/public social concerns compared with internal/personal concerns.

In comparing youth and adult perspectives, the ordering is very similar for both frequency and desirability of the ideas represented on the map. Teen, young adult, and adult sort data are strongly correlated, suggesting overall agreement among participants' conceptualizations.

NIJ

In the next phase of the study, the researchers conducted facilitated discussions with groups of adult professionals in the field and with groups of youth, including some populations not originally targeted. In adult groups, participants found the correlation surprising and encouraging, and thought the map showed the complexity of teen dating relationships in unexpected ways as well as how important cognitive and emotional map items were. In addition, the map seemed to resonate with youth. An interesting perspective emerged about how the youth saw the clusters as stages in a relationship. They also identified some aspects (such as arguing or getting many text messages) as positive or negative, depending on the context. Importantly, there was no technology cluster — social media pervades all the clusters, being integrated fully in the lives of teens.

Discussion

Following the presentation, meeting attendees raised questions about items related to sex and relationship repair. Dr. Blachman-Demner did not have the complete list of items at hand, but she noted that there were several items about sex (e.g., "having sex," "rushing into sex," and so on), although the topic didn't come up as much as expected or there wasn't much variability on it. Several items related to relationship repair (such as "getting back together" and "doing things to not create trouble") also were present. Another clarifying point was that the size of the map figures does not correspond to their importance but to their variance. Small map figures around tightly grouped items are those that most people agreed upon. Looser groups are more ambiguous.

Discussion: Improving Measurement Using the Concept Mapping Results

Dr. Blachman-Demner and Dr. Mulford continued the discussion by inviting attendees to write down their response to a guiding question. Attendees then had the opportunity to share their responses with the group. Responses are grouped here by similarity of topic where possible.

Question 3. How can the concept mapping project inform or provide a basis for future work in TDV measurement?

- The high level of correlation between teens and adults is striking. Maybe the adult lens is not unique.
- Conduct a larger scale study to see how youth vary across items.
- Work to see if it is generalizable or whether more granular work with more cultural subgroups is necessary.
- More granular work could be done for qualitative differences on the developmental level (e.g., how insecurities might differ).
- More work on populations (e.g., youth from prevention organizations who have had a lot of programming on healthy relationships compared with those who have not).
- Stages of development rather than stages of relationship — a variation?
- This map seems to represent a traditional type of relationship — what about other types (e.g., hookups, group dating)?
  - This may be the prototype or schema of teen dating relationships, not necessarily how they behave.
- Examine how this works when focused on one person's relationship compared with what teens are doing in relationships in general.

- Intervention around working with families and adolescents about what healthy dating looks like.
  - What do we want adolescent dating relationships to look like? Some youth stay coupled all the way from middle school through high school and are thought of as role models. But exploration/multiple relationships is developmentally normative.
  - We don't include the idea of the benefit of multiple relationships in our TDV prevention. We say hooking up is bad, but we don't say what is good. We don't actually know what the ideal developmental trajectory is.
  - Are we missing developmental pieces with clustering? At age 12, there's no reason to mimic adults.
- The variance of social concerns may point to a need for more measures related to peer concerns compared with family concerns.
- We talk about the variation between healthy and abusive relationships as a simple dichotomy, but the map shows layers of complexity.
- Where do youth get their ideas about relationships, and how does that modify what relationships mean? (For example, is reality TV altering perceptions?)
- Correlation with Emily Rothman's recognition of abuse scale?
- Need to validate measures of knowledge and attitude.
  - But behavior can be shifted without shifting knowledge and attitudes (e.g., kids know they're not supposed to hit someone, but they have other reasons for engaging in abuse). We need to target those reasons and situational factors (e.g., emotional regulation, impulse control, or enmeshment) that produce the abuse.
  - We need more work on cognate attitudes. People don't think it's acceptable to hit but have some cognitive representation that supports their actions. We need to get at these nuances of attitude. It could be part of a belief structure ("If you hit me, I will hit you back").
- How do friendship skills relate to risk? Part of a healthy relationship is self-regulation — if you can identify a feeling and calm down, you can act in line with your values.
  - Stability of friendships in high school is low. The changing dynamic in friendships may be informing the conceptual map. Teaching social regulation will have an impact.
  - How children are treated by parents translates to how they treat peers, which translates into how they treat dating partners.
  - But there is something important about emerging sexuality/adolescent intimacy beyond general social skills.

Discussion: Suggestions for Improving Measurement, and the Way Forward

Dr. Blachman-Demner and Dr. Mulford invited attendees to write down their responses to three more specific questions. Attendees then had the opportunity to share their responses with the group. Responses are grouped here by similarity of topic where possible.

Question 4. How can federal agencies best support the work that is needed?

Dr. Mulford urged meeting participants to be specific about funding rather than just saying we need more funding. Attendees from federal agencies did not answer this question.

NIJ

- Continue to provide funding, particularly for the existing longitudinal data sets with prospective interventions. Fund secondary data analysis of these data sets as well. We have a lot of data that has not been utilized to its full potential.
  - RFPs for scientists to propose studies/questions using existing data sets.
  - Support efforts to compile data sets with like measures to enable analysis to inform continued measurement of development.
  - A cooperative agreement approach for secondary data linkage and analysis — an addendum to grant funds for NIJ's specific goals (such as those for STRiV). Supplement existing grants like NIH/CDC does.
- Guiding integrated grant proposal solicitations to fund/guide major collaborative efforts, such as the massive efforts needed in stages for "gold standard" development.
- Funding small studies at multiple sites, addressing a specific measurement question (e.g., how do we measure dating?) and then combining data from the different studies, and having a meeting, such as a consortium.
  - This could include training in new ways of data collection.
  - Variations: comparative agreement or replication studies.
  - Examining cultural/high-risk/other subgroups between sites.
- Writing grant announcements that require justification for use of a particular measure. Set measurement standards like journals do.
  - More emphasis on the measurement section and more explicit guidance would help build in more value.
  - Consider announcements specifically for measurement — to improve, develop, test and experiment with measures (including testing with diverse groups).
  - Identify the measure in the announcement to be used, and make funding conditional on the use of that measure.
- In general, writing more targeted and specific RFPs. Provide funding to answer specific research questions.
  - Consider cooperative agreements instead of preparing proposals for giant pools of applicants, where the odds of funding are low. Use funds for theme issues of journals, organizing working meetings that result in products, but most importantly, provide funding to accomplish a very narrowly specified part of the work (e.g., we do not need more etiology studies in violence).
- Being thoughtful and deliberate in choosing review panels.
- Investment by the federal government in knowledge management/synthesis/evaluation (informatics) to help avoid reinventing the wheel.
  - Studies-based analysis could potentially help use research to inform policy. It is a novel idea that helps identify gaps in what has been studied already.
  - A systematic scan of what we know and what we don't.
  - Create and facilitate a "live" online community for discussion/sharing (i.e., a targeted listserv with or without a structured webpage of current best practices). An online community (TreatTobacco.net) is an example.
  - Training on using new technology for data collection and intervention.

- Fund more cross-disciplinary or cross-topic research, possibly independent of adolescent/developmental research (e.g., domestic violence, child abuse, adolescent relationships, and peer groups).
- Fund more cross-project collaborations where leading researchers work together and use their expertise to tackle these ideas rather than having individuals working in their own silos.
  - Foster multi-university collaborations to increase sampling representativeness and efficiencies in data collection efforts.
- Convene work groups with facilitator/support staff; organize conference calls and in-person meetings with these groups.
  - Planning workshop series with specific product outcomes in mind to advance the field/research (e.g., compendium of data sources, gold standard approaches, considerations from other fields/applications in advancing the rigor of methods, stages of tech intervention development).
  - Electronic ways to connect researchers over time as a group.
- A specific RFA to fund meta-analysis.
- A methodology study solicitation.
- More emphasis on measures of peer norms and behaviors, family processes, and more appropriate attitude scales rather than actual measures of TDV.
  - Enhanced measurement to all domains that predict the D.V. — norms, attitudes toward use of violence, link to emotions, neighborhood context, family factors — we need added specificity in these areas. The D.V., whether it's the CTS-2 or others, performs well and most studies devolve into a yes/no — what is needed is screening for youth on the high/serious end (e.g., Campbell's risk for homicide scale).
- More qualitatively based research to better understand different cultural subgroups to see whether current measures can be applied to them or what would do better.
  - An add-on to develop or validate measures across subgroups or cultures.
- Surveillance measures and standardized data collection.
- Translate research into policy and practice.
- Invest in knowledge management and encourage innovation.
- Strategically allocate funding and carefully declare things — everyone wants their own agenda to be the agenda. At some point, imposed structure will facilitate progress.

Question 5. How useful would a document be (e.g., an addendum to the CDC IPV compendium of measures) that provides guidance for measuring TDV and specifically focuses on developmentally appropriate questions for different age groups?

Attendees were asked to rate the usefulness of such a document on a four-point scale. In an informal show of hands, about 15 attendees rated it as "very useful," 2 rated it as "somewhat useful," 1 rated it as "not very useful" and no attendees raised their hands for the "not at all useful" category. Then attendees shared the reasons for their selections with the group. Responses are grouped here by similarity of topic.

- Very useful.
  - Researchers may be more likely to include TDV measures in their studies if they are easily accessible and if there are short items or a few key items that could be highlighted.

NIJ

- It may help improve matching of questions to sample populations across the field; need to consider age ranges and comparability carefully. It would be helpful to know what had been validated and with what age group, etc.
- Authors of the scales should be invited to write up their intent with the measure and concerns with implementing their scales (e.g., pitfalls to avoid). Also, add studies that used the scales in treatment studies and basic research.
- In addition to easily accessing measures that would allow comparisons among studies, we also need good screening tools.
- It would be very useful but time consuming — consider an RFP or a cooperative agreement with an existing grantee to compile this.

- Drawbacks of a book.
  - Could become outdated quickly.
  - Focused on initial publication only.
  - Something weak could be put right next to something nationally known without noting the difference. Frontline practitioners want a ranking, not just a long list — they need more guidance.
  - This would not be useful — we need to evaluate and then eliminate.
  - It would be useful, to the extent it reflects valuable resources, but should not become prescriptive (actually or in perception) where it would be at cross-purposes and stifle new work on measures/adaptations because of the perceptions that these measures should be used to get funding.
  - The existing models/measures are what they are, and most of us know them and can find them. I'm not sure collecting them would add much; however, careful, systematic construct validation and subsequent developmental measurement research would be helpful — but this is MUCH more expensive.
  - It could be useful but the bulk of the work in this area is using secondary data and so we need to essentially deal with the limitations of existing measures. It seems that it's just as important to increase our understanding of the problems of current measurement strategies as developing new ones.

- Useful alternatives.
  - A Wikipedia-like page that we could update and exchange ideas about what worked.
  - A live addendum to the IPV measures compendium would be constructive. Working group efforts to update the compendium for publication as a static effort is not likely to stay on the front edge even until publication. So a tracked/working history as discussion develops to allow for consistency in longitudinal work as well as transitioning/testing new measures.

- But hard copies for practitioners. They like to have books.
  - Following the National Institute on Alcohol Abuse and Alcoholism's Clinician's Guide model, to give recommendations based on specific criteria ("if you can only put in n items, do this, and here's the evidence to back it up"). Very helpful, although a lot of work.
  - Giving ratings (as *Consumer Reports* does) or at least incorporating feedback in registries of programs.

- Not enough research yet to be able to rate scales or fund research to enable comparisons.

  Need to start with large item pool and test these items with different groups by age, gender, ethnicity, state of relationship and geography. This would be useful to adapt items/measures

for specific populations rather than using a one-size-fits-all. This would, however, require substantial empirical testing of psychometric properties but ideally could result in weights of item parameters (loadings, thresholds) that could be useful for scoring in smaller samples — but these parameters need to be focused and developed for different populations or be nationally representative.

- Need to provide information on scoring items to increase uniformity.

Question 6. How helpful would it be to develop a gold standard for measurement of TDV as Follingstad and Bush (2014) propose for IPV?

Attendees were asked to rate the helpfulness of developing a gold standard on a four-point scale. In an informal show of hands, about 2 attendees rated it as "very helpful," 6 rated it as "somewhat helpful," 7 rated it as "not very helpful," and no attendees raised their hands for the "not at all helpful" category. Group discussion followed. Responses are grouped here by similarity of topic.

- Process of establishing a gold standard.
  - Something isn't a gold standard because we say it is; it's because we demonstrate that it's the gold standard, which is a lot of work. Also, just because it is created doesn't mean everyone will use it.
  - The process may be very long and, at the end, the measure could be outdated.
  - It would be very expensive because it would need to be tested on so many populations. If that is not done, then it would be difficult to know how relevant the gold standard is to specific groups.
  - There should also be room for other measures to be used as a supplement or tested because the gold standard may not be sufficient for all participant groups or purposes.
  - We could have several standards for different research purposes.
  - Many dimensions of TDV/teen relationships require in-depth study (including subgroups of youth), and it's not going to result in one standard. An integrated, parallel development process would be the most constructive.
- Establishing criteria for a gold standard.
  - One fixed measure might not be helpful, but creating criteria and processes for evaluating that go beyond the current level of science would be helpful.
  - Agreeing upon a "gold standard" set of requirements for choosing the measure you choose would be useful. Maybe a list of what needs to be reported (e.g., does the measure function well in the population of interest, for the purpose you are using it (screening, intervention, prevalence)? What are the reasons you selected *x* measure?
  - Need to decide how important it is to use agreed-upon definitions, wording, and preambles.
  - We do need measures for different purposes and to support individual research groups in creating them (e.g., screening instruments, instruments sensitive to intervention effects, instruments for particular ethnic groups). Agreeing upon these different types of measures could be useful.
  - Having a fixed measure could constrain innovation and science. Instead, if a measure could have some sort of rating (such as *Consumer Reports*), researchers could see criteria on it that might include how much literature base it has, its validity, multimethod

convergence, etc. Guidelines for separate criteria could encourage researchers to evaluate them, and this would help them think what excellence in measurement means.

– There is not one gold standard per se; it depends on what purpose you want to measure violence for (e.g., capture low-level violence or not?). Better to come up with a gold set of standards of criteria you have to meet.

– It depends on what kind of work you are doing and what your question is. If you are doing prevention/intervention work, you want to be very sensitive to assessing change (e.g., daily electronic assessment might be very sensitive).

– We need to allow room for growth — flexibility in the research approach as technology advances, youth culture changes, definitions are updated, etc.

– We don't need to start from scratch, especially if we do decide to develop a gold standard and/or criteria.

- Suggestions.

  – If going in the direction of having a Wikipedia-like page, maybe frame it around what we don't know (where the gaps are) instead of where a measure has been used.

  – A list of different dimensions and rating the importance of the dimensions (e.g., content areas covered, frequency of administration, and so on) would help show the full scope of dimensions that are relevant and which ones further our understanding.

  – When creating a matrix of what has and hasn't been done, consider making it broad and include adult populations. Think about what has been done with adults but not teens.

  – The NIH PhenX Toolkit and PROMIS Measures might be useful to look at.

Recommendations for Next Steps in TDV Measurement, and Wrap-Up

Dr. Mulford thanked meeting attendees for their participation. She listed a few concrete steps that NIJ would likely take:

- Look at funding research on measurement, especially related to those issues that can move the field forward.
- Create a consortium to bring researchers together on a measurement-related topic.
- Require applicants for TDV research to justify the use of a measure.

She also listed some actions that NIJ would like to do:

- Establish criteria for measures — backing away from the idea of a gold standard, but moving toward a clearinghouse that includes adult IPV and TDV as a cross-topic.
- When addressing questions from the field about what measures to use, have a more standardized way of making recommendations that would be dynamic (e.g., if you have space for a few items, use this, but if you have space for more items, use that; if working in *x* setting, use this, but if in *y* setting, use that).

A final comment was raised that there are other people doing work on this topic who were not part of this meeting. NIJ wants to be aware of investigators doing relevant work.

For those not in attendance, notes will be made public.

NIJ