Interpreting the Science: Translating Systematic Reviews into Guidelines

Shawna L. Mercer, MSc, PhD, Holly Wethington, PhD, Leigh Ramsey Buchanan, PhD, Ramona K. Finnie, DrPH, Devon Okasako-Schmucker, MPH

Workshop of the Committee to Review the Process to Update the Dietary Guidelines for Americans
National Academies of Sciences, Engineering, and Medicine
January 10, 2017        Washington, DC
Disclaimer

The findings and conclusions in this presentation are those of the presenters and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

The Centers for Disease Control and Prevention provides “ongoing administrative, research, and technical support for the operations of the Task Force.”

(PhS Act § 399U(c))
Agenda

- Introduction to The Guide to Community Preventive Services (The Community Guide)
- Translating systematic review findings into Community Preventive Services Task Force recommendations
  - The component parts
  - Putting it all together
- Decisions about updating Community Guide systematic reviews
Introduction to The Community Guide
State-of-the-science systematic reviews that

• Analyze all available evidence on the effectiveness of population-based programs, services, and policies (interventions) in
  o Communities
  o Health systems

• Assess the economics of all effective interventions

• Highlight critical evidence gaps
Evidence-based recommendations developed by the Community Preventive Services Task Force based on the systematic review findings

www.thecommunityguide.org
Community Preventive Services Task Force (Task Force)

- A non-federal, independent, rotating panel
- Internationally renowned experts in public health research, practice, and policy
- Nomination process includes broad input from throughout public health and healthcare
- Members are appointed by the CDC Director
- Serve without payment

  - CDC provides scientific, technical, and administrative support for the Task Force
# Topics for Task Force Reviews (as of 2016)

<table>
<thead>
<tr>
<th>Reviews Organized by Environment</th>
<th>Health equity (Social Environment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews by Risk Behavior</td>
<td>Reviews by Specific Condition</td>
</tr>
<tr>
<td>Alcohol abuse/misuse</td>
<td>Cancer</td>
</tr>
<tr>
<td>Tobacco use</td>
<td>Mental health</td>
</tr>
<tr>
<td>Poor nutrition and obesity</td>
<td>Vaccine-preventable diseases</td>
</tr>
<tr>
<td>Physical inactivity</td>
<td>Violence</td>
</tr>
<tr>
<td>Unhealthy sexual behaviors</td>
<td>Motor vehicle injuries</td>
</tr>
<tr>
<td>Cardiovascular disease prevention</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Orca lhealth</td>
<td>Oral health</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reviews Organized by Setting</th>
<th>Reviews Organized by Life Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worksite health promotion</td>
<td>Adolescent health</td>
</tr>
<tr>
<td>Special Projects</td>
<td></td>
</tr>
<tr>
<td>Health communication</td>
<td>Emergency preparedness</td>
</tr>
</tbody>
</table>
Translating Systematic Review Findings into Task Force Recommendations: The Component Parts
Critical Questions to Address About a Population-Based Public Health Intervention

- Does it work?
- How well?
- For whom?
- Under what conditions?
- How does it influence health disparities?
- What is the cost?
- Does it provide value?
- What are important considerations for implementation?

Tentative or partial answers are preferable to no answers (especially for the latter questions)
In General, a Conclusion About Effectiveness Requires....

**A Body of Evidence**
- Number of studies
- Quality of studies
  - Study limitations
- Suitability of study design

**A Demonstration of Effectiveness**
- Consistency of Effect
  - “Most” studies demonstrated an effect in the direction of the intervention
- Sufficient Magnitude of Effect
  - The effect demonstrated across the body of evidence is “meaningful”
Task Force Recommendation Options

- **Recommend against**
  - Strong Evidence
  - Sufficient Evidence

- **Insufficient evidence to recommend for or against**

- **Recommend**
  - Strong Evidence
  - Sufficient Evidence
Typical Community Guide Review Coordination Team

- 1 coordinating scientist
- 1–2 research fellows
- 3–6 subject matter experts
- 1–2 Task Force members
- 1–2 Task Force Liaison members
Conceptualization of the Systematic Review

- Draft four components
  - Intervention definition
  - Research questions
  - Analytic framework
  - Applicability factors

- Dynamic process with constant team input to determine the appropriate level of breadth and depth for usefulness to the field
Conceptualize: Intervention Definition

- A combination of definition and description of the intervention
  - Wording:
    - Terms commonly used in the field
    - Understandable by the public
  - Distinguishes between
    - Essential ("must haves")
    - Non-essential but still important ("may have") intervention components

- Intervention definition evolves over the course of a review
Conceptualize: Research Questions

- Primary research questions
  • How effective is the intervention in producing the desired outcomes?

- Secondary research questions
  • Does the intervention effectiveness change because of characteristics of the
    o Population?
    o Setting?
    o Intervention?
Analytic Framework: Behavioral Interventions that Aim to Reduce Recreational, Sedentary Screen Time

Potential Harms: None identified

Additional Benefits:
- Educational activities
- Social interaction
- Reduced exposure to violent or adult content
- Improved sleep

Key Effect Modifiers:
- Intervention Intensity
- Population Characteristics: age, SES, and weight status
- Peer Influence

Increased:
- Knowledge
- Awareness
- Attitude
- Motivation
- Skills

Reduced access:
- TV
- Video games
- Computers

Reduced exposure to unhealthy food advertising

Increased Physical Activity

Reduced Sedentary Screen Time

Improved Diet (e.g., unhealthy snacking)

Improved Weight-related outcome (e.g., weight, BMI, % body fat)

Reduced Morbidity & Mortality

Reduced Disparities

Improved Diet

Increased Physical Activity

Reduced Sedentary Screen Time

Improved Weight-related outcome (e.g., weight, BMI, % body fat)

Reduced Morbidity & Mortality

Reduced Disparities

Improved Diet

Increased Physical Activity

Reduced Sedentary Screen Time

Improved Weight-related outcome (e.g., weight, BMI, % body fat)

Reduced Morbidity & Mortality

Reduced Disparities

Improved Diet

Increased Physical Activity

Reduced Sedentary Screen Time

Improved Weight-related outcome (e.g., weight, BMI, % body fat)

Reduced Morbidity & Mortality

Reduced Disparities

Improved Diet

Increased Physical Activity

Reduced Sedentary Screen Time

Improved Weight-related outcome (e.g., weight, BMI, % body fat)

Reduced Morbidity & Mortality

Reduced Disparities
Body of Evidence: Abstracting Study Quality

- “Study quality” is sometimes misunderstood

- We are interested in the utility of the study for helping to answer our research questions

- A study that is very well-done may not be rated as high quality because:
  - There are confounding factors beyond the researcher’s control
  - The researcher was attempting to answer a different question than we were
Quality of Study Execution

- To assess limitations to external validity and internal validity
- Nine possible limitations for a given study (across six domains)
- Each team makes additional, review-specific decision rules to assess study quality
# Quality of Execution Assessment Framework

<table>
<thead>
<tr>
<th>Domain</th>
<th>Potential Reasons for Limitations</th>
<th>Maximum Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>• Was the study population well described?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• etc.</td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>• Was the sampling frame/universe adequately described?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• etc.</td>
<td></td>
</tr>
<tr>
<td>Measurement</td>
<td>• Were outcome measures valid and reliable?</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>• etc.</td>
<td></td>
</tr>
</tbody>
</table>
### Quality of Execution Assessment Framework (cont’d)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Potential Reasons for Limitations</th>
<th>Maximum Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Analysis</td>
<td>• Appropriate statistical testing conducted?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• etc.</td>
<td></td>
</tr>
<tr>
<td>Interpretation of Results</td>
<td>• &gt;80% completion rate? Data set complete?</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>• etc.</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>• Other biases or concerns not included in the previous domains (e.g., evidence of selective reporting)</td>
<td>1</td>
</tr>
</tbody>
</table>
Quality of Execution Categories

• 0–1 limitations = Good Quality

• 2–4 limitations = Moderate Quality

• 5+ limitations = Limited Quality (Excluded)
Body of Evidence: Suitability of Study Design

- Greatest suitability
  - RCT (individual or group)
  - Non-Randomized Trial
  - Prospective Cohort
  - Other Design With Concurrent Comparison

- Moderate suitability
  - Interrupted Time Series
  - Retrospective Cohort
  - Case-Control

- Least suitability
  - Uncontrolled Before-After
  - Cross-Sectional
Challenge: Assessing Effectiveness with Multiple Outcomes

Researchers may agree on the important intended outcomes (e.g., increased food and vegetable intake), but

1) There may be considerable variability in how outcomes are measured

   e.g., studies variably report food and vegetable intake in

   - Servings/day
   - Grams/day
   - Odds ratio
   - Percent of energy intake
Challenge: Assessing Effectiveness with Multiple Outcomes (cont’d)

2) Outcomes of interest may be combined and not able to be disentangled
   e.g.,
   - 100% fruit juice and sugar sweetened beverages
   - Low fat and high fat dairy products

- Lack of consensus among researchers on favorable direction of dietary outcomes
  e.g.,
  - Total energy intake among children
  - Dairy intake
  - 100% fruit juice intake
Assessing Effectiveness with Multiple Outcomes

- The Community Guide identifies the range of measures of outcome
- It attempts to ‘lump’ sufficiently related outcomes together
- Direction and magnitude of effects are presented when quantitative summaries are not possible
- **But:** there remain considerable implications in terms of loss of power, often leading to a finding of ‘insufficient evidence’ to recommend for or against an intervention
Typical Evidence Presentation: Behavioral Screen Time Review
Evidence on Effectiveness – Commercial Television Viewing

Randomized Controlled Trial

Before/After

n=7 studies
(9 arms)

HIGH INTENSITY
(6 studies)
Median: -0.78
hours/day
IQI: (-3.4, -0.61
hours/day)

LOW INTENSITY
(4 studies)
Median: -0.26
hours/day
(-15.6 minutes/day)
Range (-2.0, 0
hours/day)

High
Intensity (Above Solid Horizontal Line)
Robinson 1999*, † (7m) (2.2)
Robinson 2006*, † (6m) (2.2)
Zimmerman 2012 (4m) (0.9)
Ni Mhurchu 2009 TVMGR ARM § (1.5m) (1.9)

Gorin 2006 *, † (2m) (7.5)
Ford 2002 TVMGR ARM §, § (1m) (7.7)

Low
Intensity (Below Solid Horizontal Line)
Escobar-Chavez 2010 (6m) (5.8)
Ni Mhurchu 2009 ADVICE ARM § (1.5m) (1.2)
Ford 2002 ADVICE ARM ‡, § (1m) (5.6)
Robinson 2006 SIBLING*, † (6m) (1.7)
Robinson 2006 DAD*, † (6m) (1.6)
Robinson 2006 MOM*, † (6m) (1.5)

LEGEND
* Significant † Adjusted change ‡ Family measure § Pre-post study arm
IQI, Interquartile Interval

Absolute Change in Commercial TV Viewing (hours/day)
Typical Evidence Presentation: Behavioral Screen Time Review
Evidence on Effectiveness – Changes in Fruit and Vegetable Intake

**Study (Follow up) (Intervention group baseline servings/day)**

- **High Intensity**
  - **Youth + Adults**: Median: 0.21 servings/day (IQI: 0.12, 0.57 servings/day)
  - **Youth** (8 studies, 10 arms): Median: 0.16 servings/day (IQI: 0.25, 0.43 servings/day)
  - **Adults** (2 studies, 3 arms): Median: 0.60 servings/day (range: 0.11, 4.3 servings/day)

- **Low Intensity**
  - **Youth**: Median: 0.50 servings/day (IQI: 0.85, 1.04 servings/day)
  - **Favors Intervention**

**Absolute Change in Fruit and Vegetable Intake (servings/day)**

- **Warren 2003 PA+Diet (14 m)** (1.76)
- **Whaley 2010 (12 m)** (5.6)
- **Gentile 2009 (6m)** (4.5)
- **Branscum 2011 SCT (1m)** (3.41)
- **Branscum 2011 K (1m)** (3.35)

**LEGEND**
- *Significant effect
- ¶ Adjusted change
- § Pre-post study arm
- **Sig effect for girls (p=0.003), but not boys
- Bolded studies are overweight/obese samples
- IQI, Interquartile Interval
**Typical Analyses and Presentation Decisions: Legend for Tables**

<table>
<thead>
<tr>
<th>Direction and Significance</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable direction, statistically significant</td>
<td>++</td>
</tr>
<tr>
<td>Favorable direction, approaching significance (0.05&lt;p&lt;0.10) or statistical significance not calculable</td>
<td>+</td>
</tr>
<tr>
<td>Null effect</td>
<td>Ø</td>
</tr>
<tr>
<td>Unfavorable direction, statistically significant or not</td>
<td>-</td>
</tr>
</tbody>
</table>
## Typical Evidence Presentation: Behavioral Screen Time Review

### Evidence on Effectiveness: Changes in Energy Intake

<table>
<thead>
<tr>
<th>Study (Location) Study Design</th>
<th>Study Population (n)</th>
<th>Outcome</th>
<th>Length of Follow-up</th>
<th>Intervention Baseline</th>
<th>Finding (p-value or 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Screen Time Intensity: High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Otten '09 (Chittenden County, VT)</td>
<td>Obese adults ages 21-65 viewing ≥3 hrs/d TV (n=36)</td>
<td>Energy intake</td>
<td>1.5 mos</td>
<td>2299.7 kcal/d</td>
<td>-87 kcal/d (0.52)</td>
</tr>
<tr>
<td>Epstein '08 (Buffalo, NY)</td>
<td>Children ages 4-7 with BMI at or above 75th percentile (n=70)</td>
<td>Energy Intake</td>
<td>6 mos</td>
<td>1551.4 kcal/d</td>
<td>-75 kcal/d (NS)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 mos</td>
<td></td>
<td></td>
<td>-150 kcal/d (0.05)</td>
</tr>
<tr>
<td>Ni Mhurchu '09 (New Zealand) Advice + TV Manager</td>
<td>Children ages 9-12 viewing TV &gt;20 hrs/wk (n=15)</td>
<td>Energy intake from snacks</td>
<td>1.5 mos</td>
<td>767.1 kcal/d</td>
<td>-233.6 kcal/d (NR)</td>
</tr>
<tr>
<td><strong>Screen Time Intensity: Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni Mhurchu '09 (New Zealand) Advice Only</td>
<td>Children ages 9-12 viewing TV &gt;20 hrs/wk (n=14)</td>
<td>Energy intake from snacks</td>
<td>1.5 mos</td>
<td>689.7 kcal/d</td>
<td>-123.7 kcal/d (NR)</td>
</tr>
</tbody>
</table>
### Typical Evidence Presentation: Behavioral Screen Time Review Evidence on Effectiveness: Changes in Fruit and Vegetable Intake

<table>
<thead>
<tr>
<th>Study (Location) Study Design</th>
<th>Study Population (n)</th>
<th>Outcome</th>
<th>Length of Follow-up</th>
<th>Intervention Baseline</th>
<th>Finding (p-value or 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Screen Time Intensity: High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marcus 2009 (Sweden)</td>
<td>6-10 yrs/Grades 1-4 (n=692)</td>
<td>Fruit and Vegetable intake</td>
<td>48</td>
<td>NR</td>
<td>0.025 difference in proportions (NS)</td>
</tr>
<tr>
<td>de Silva-Sanigorski 2010 (Australia)</td>
<td>0-5 yrs (n=993–1,022)</td>
<td>Vegetable intake</td>
<td>36</td>
<td>1.07 previous d</td>
<td>Regression coefficient 0.10 (-0.01, 0.20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fruit intake</td>
<td></td>
<td>1.29 previous d</td>
<td>Regression coefficient 0.07 (-0.02, 0.16)</td>
</tr>
<tr>
<td>Gortmaker 1999 EWKM (Baltimore, MD)</td>
<td>Grades 4-5 (n=336)</td>
<td>Fruit and Vegetable intake</td>
<td>20</td>
<td>1.6 per 1000 kcal</td>
<td>0.13 per 1000 kcal* (-0.04, 0.30)</td>
</tr>
</tbody>
</table>

**Screen Time Intensity: Low**

<table>
<thead>
<tr>
<th>Study (Location) Study Design</th>
<th>Study Population (n)</th>
<th>Outcome</th>
<th>Length of Follow-up</th>
<th>Intervention Baseline</th>
<th>Finding (p-value or 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campbell 2013 (Australia)</td>
<td>Newborns (recruited first-time parents) (n=480)</td>
<td>Fruit intake (24 hr recall)</td>
<td>15</td>
<td>NR</td>
<td>13.33 g/d (-2.59, 29.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vegetable intake (24 hr recall)</td>
<td></td>
<td>NR</td>
<td>6.62 g/d (-2.51, 15.76)</td>
</tr>
</tbody>
</table>
Alternate Analysis Presentation: Built Environment Interventions

Summary Results Legend

Cross-sectional Associations

++  Favorable direction, statistically significant

+   Favorable direction, not statistically significant or testing not reported (NR)

Ø   Null association

-   Unfavorable direction, not statistically significant or testing NR

+-  +/- at least one outcome measure in each direction (favorable + unfavorable)
## Alternate Analysis Presentation: Built Environment Interventions

<table>
<thead>
<tr>
<th>Study</th>
<th>Transport walk/bike</th>
<th>Recreation walk/bike</th>
<th>Total Walking</th>
<th>Total Physical Activity</th>
<th>MVPA</th>
<th>Meets Guideline Recommended MVPA</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maddison 2009</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sallis 2009</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Dyck 2009</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zenk 2009*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Dygryn 2010</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norman 2010 (Girls)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norman 2010 (Boys)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graham 2011</td>
<td></td>
<td></td>
<td>Ø</td>
<td>Ø</td>
<td></td>
<td>Ø</td>
<td></td>
</tr>
<tr>
<td>Sundquist 2011</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carlson 2012</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ding 2012</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ding 2012</td>
<td>+</td>
<td>++</td>
<td>Ø</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDonald 2012</td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robertson 2012</td>
<td></td>
<td></td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sayers 2012</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whitten 2012</td>
<td>Ø</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeMeester 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Self-reported adherence to walking prescription*
Challenge: Addressing the Needs of Decision Makers

- Always need to assess whether the intervention works under the conditions set forth in the study
  - Internal validity

- But most decision makers considering community preventive services want to know
  - Are the findings generalizable across all the settings, situations, and populations for which I am responsible?
  - Are the findings applicable to my specific setting, situation, or population?
Applicability Considerations

- At the start of each review
  - What applicability factors are relevant?
  - What are our a priori hypotheses?

- During the review
  - Collect data from included studies to answer each a priori hypothesis

- In the concluding phase of the review
  - Can we confirm/reject our a priori hypotheses?
  - What can we conclude? Is the intervention
    - Effective across settings and population groups?
    - Only effective in certain environments?
Conceptualization: Applicability Considerations

- We routinely consider
  - Settings
  - Population characteristics
  - Intervention characteristics

- We always consider
  - U.S. vs. non-U.S.
  - Race/ethnicity
  - Socioeconomic status: income, education

- Each review is unique; the team asks
  - Is there evidence suggesting a particular factor needs to be considered?
  - Are any normally considered factors irrelevant to this intervention?
Applicability Considerations: Sources of Information

- **Empirical evidence**
  - Included studies reported on one or more of the applicability factors under consideration
  - Main source of information

- **Broader literature**
  - Other reviews, opinion pieces, commentaries, and reports relevant to the intervention

- **Expert opinion**
  - Subject matter experts on review coordination teams, consultants

- **Theoretical considerations**
### Example: Coordination Team Initial Discussions on Applicability
Behavioral Interventions that Aim to Reduce Recreational, Sedentary Screen Time

<table>
<thead>
<tr>
<th>Factor(s)</th>
<th>How is this Factor Related to Intervention Effectiveness?</th>
<th>Points from Coordination Team Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>US/Non-US</td>
<td>Unsure</td>
<td>Non-US may not be comparable to US due to various advertising bans</td>
</tr>
<tr>
<td>Urban/Suburban/Rural</td>
<td>Unsure</td>
<td>Urban (safety) and rural (access) settings likely to have less options</td>
</tr>
<tr>
<td>Community/Worksite/Healthcare/School/Faith-based</td>
<td>Probably applicable</td>
<td>Likely applicable in all settings</td>
</tr>
<tr>
<td>Age</td>
<td>Probably effect modifier</td>
<td>Effect is likely to be greater among children than for adolescents due to parental control</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>Probably applicable</td>
<td>Likely applicable for participants across racial and ethnic groups</td>
</tr>
<tr>
<td>SES</td>
<td>Probably effect modifier</td>
<td>Effect is likely to be smaller in magnitude among low-SES due to fewer options for alternative pastimes</td>
</tr>
<tr>
<td>Weight Status</td>
<td>Probably effect modifier</td>
<td>Effect is likely to be smaller in normal-weight population due to floor effect</td>
</tr>
</tbody>
</table>
Translating Community Guide
Systematic Review Findings into Task Force Recommendations:
Putting It All Together
From Evidence to Task Force Recommendations: Putting It All Together

- **Primary consideration for Task Force Finding Statement**
  - Intervention effectiveness
  - Consistency and meaningfulness of results
    - Number of included studies
    - Study design

- **Additional factors considered for Task Force Finding and Rationale Statement**
  - Applicability
  - Other Effects
    - Harms? Do harms outweigh intervention benefits?
    - Additional benefits?
  - Considerations for Implementation
  - Evidence Gaps
In General, a Conclusion About Effectiveness Requires….

A Body of Evidence
- Number of studies
- Quality of studies
- Study limitations
- Suitability of study design

A Demonstration of Effectiveness

Consistency of Effect
“Most” studies demonstrated an effect in the direction of the intervention

Sufficient Magnitude of Effect
The effect demonstrated across the body of evidence is “meaningful”
Interpretation of Effect Sizes: Meaningful Effect

- Task Force looks for effectiveness in achieving changes that is meaningful in a public health or population-based context

- This usually requires a judgment based on available (but often limited) information
  - Evidence for and interpretation of meaningfulness may differ across topics and the available evidence
  - Factors considered include
    - The intervention
    - Study designs and comparisons
    - The evaluated outcome and location on the causal pathway
    - The measurement used to determine change
Interpretation of Effect Sizes: Meaningful Effect

- Task Force judgement may be informed by effect sizes considered clinically meaningful within that topic
  - Especially for intervention reviews with similar causal pathways and outcomes

- Coordination Team and Task Force assessments may also include
  - Information from the broader literature
  - Effect sizes from previous intervention reviews (in the same topic)
  - Coordination Team discussion
  - Task Force discussion
## Community Guide Translation Table

<table>
<thead>
<tr>
<th>Evidence of Effectiveness</th>
<th>Quality of Execution</th>
<th>Suitability of Design</th>
<th>Number of Studies</th>
<th>Consistent</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRONG</td>
<td>Good</td>
<td>Greatest</td>
<td>2 or more</td>
<td>Yes</td>
<td>Meaningful</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>Greatest or Moderate</td>
<td>5 or more</td>
<td>Yes</td>
<td>Meaningful</td>
</tr>
<tr>
<td></td>
<td>Good or Fair</td>
<td>Greatest</td>
<td>5 or more</td>
<td>Yes</td>
<td>Meaningful</td>
</tr>
<tr>
<td></td>
<td>Meet criteria for SUFFICIENT but not STRONG body of evidence</td>
<td>LARGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUFFICIENT</td>
<td>Good</td>
<td>Greatest</td>
<td>1</td>
<td>NA</td>
<td>Meaningful</td>
</tr>
<tr>
<td></td>
<td>Good or Fair</td>
<td>Greatest or Moderate</td>
<td>3 or more</td>
<td>Yes</td>
<td>Meaningful</td>
</tr>
<tr>
<td></td>
<td>Good or Fair</td>
<td>Greatest Moderate Least</td>
<td>5 or more</td>
<td>Yes</td>
<td>Meaningful</td>
</tr>
<tr>
<td></td>
<td>Meet criteria for STRONG body of evidence, but Task Force opts to downgrade conclusion to SUFFICIENT for one or more reasons</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert Opinion</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
<td>Meaningful</td>
</tr>
<tr>
<td>INSUFFICIENT (one or more)</td>
<td>Inadequate designs or execution</td>
<td>Too Few</td>
<td>No</td>
<td>Small</td>
<td></td>
</tr>
</tbody>
</table>
Example Overall Review Assessment: Behavioral Interventions that Aim to Reduce Recreational, Sedentary Screen Time

- **Body of evidence:** 12 studies with 14 study arms (Screen-Time-Only)
  - 8 studies with greatest suitability of design
  - 67% RCTs or group RCTs
  (See translation table)

- **Impact on screen time (12 studies, 14 study arms)**
  - Magnitude of effect meaningful? Meaningful (substantial)
  - Consistent across the body of evidence? Consistent

- **Impact on weight-related outcomes (6 studies, 7 study arms)**
  - Magnitude of effect meaningful? Meaningful (modest)
  - Consistent across the body of evidence? Consistent

- **Impact on physical activity (6 studies, 8 study arms)**
  - Magnitude of effect meaningful? Meaningful (modest)
  - Consistent across the body of evidence? Consistent

- **Impact on diet (9 studies, 11 study arms)**
  - Magnitude of effect meaningful? Meaningful (substantial)
  - Consistent across the body of evidence? Consistent

- **Impact on disparities**
  - Greater effectiveness in lower SES and lower education

- **Coordination Team assessment**
  - Strong evidence of effectiveness
Current Options for Qualifying an Intervention Review

Conclusion

Sufficient Body of Evidence (Recommend)

Downgrade to Recommend Against
- Major harm(s)

Upgrade to Recommended (strong)

Downgrade to Insufficient Evidence
- Serious recurring flaw(s)
- Concerns about link to health outcome(s)
- Magnitude concerns
- Applicability concerns
- Harm(s) concerns

Narrowed Recommendation
- Applicability concerns
- Differential harm

Narrative Qualification
- Significant barriers
- Other concerns
  - Intervention
  - Evidence
Role of Judgment in the Community Guide Process

- Community Guide methods are systematic and transparent
- Nonetheless, judgments are needed at multiple stages
- Broad participation helps to inform those judgments and reduce bias
  - Systematic review experts
  - Subject matter researchers
  - Subject matter practitioners
  - Community Preventive Services Task Force members
  - Task Force Liaison members
Deciding When to Update Community Guide
Systematic Reviews
The Task Force’s Work List

Updates Priority List
1. e.g., Physical Activity
2. ...
3. ...
4. ...
5. ...

New Topics Priority List
1. e.g., Obesity and Nutrition
2. ...
3. ...
4. ...
5. ...
6. ...

Special Projects List
1. e.g., H1N1
2. ...
3. ...
4. ...
5. ...

Work List
A combination of the categories
# Task Force-Approved Criteria for Prioritization of Community Guide Review Topics

1. Burden
2. Preventability
3. Important disparities
4. Ability to cover reasonable portion of topic
5. Balance across public health topics
6. Audience/stakeholder interest
7. Alignment with other national efforts
8. The degree to which Task Force work can make a difference

<table>
<thead>
<tr>
<th>Dimension/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data and Epidemiology</td>
</tr>
<tr>
<td>Review Content and Interconnections</td>
</tr>
<tr>
<td>Review Environment and Support</td>
</tr>
<tr>
<td>Approved Task Force (TF) criteria</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Burden</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Preventability</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Presence of important disparities</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Ability to cover reasonable portion of topic within a 2 to 4 year project period</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Balance across public health topics considered by the TF</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Audience/Stakeholder interest in Task Force work in this topic</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alignment with other federal or national efforts</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>The degree to which the Task Force can make a difference</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
THANK YOU!

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.
Scope of Community Preventive Services Task Force (CPSTF) and US Preventive Services Task Force (USPSTF)

## Icons in *Community Guide* Analytic Frameworks

<table>
<thead>
<tr>
<th>Icon</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Circle" /></td>
<td>Intervention</td>
</tr>
<tr>
<td><img src="image" alt="Box" /></td>
<td>Recommendation outcome</td>
</tr>
<tr>
<td><img src="image" alt="Box" /></td>
<td>Other intermediate outcome/variable (that are not recommendation outcomes)</td>
</tr>
<tr>
<td><img src="image" alt="Population" /></td>
<td>Population</td>
</tr>
<tr>
<td><img src="image" alt="Key Effect Modifiers" /></td>
<td>Key Effect Modifiers (affecting causal relationships)</td>
</tr>
<tr>
<td><img src="image" alt="Additional Benefits/ Potential Harms/ Disparities" /></td>
<td>Additional benefits/Potential Harms/Disparities</td>
</tr>
<tr>
<td><img src="image" alt="Unidirectional Block Arrow" /></td>
<td>Unidirectional block arrows are applied between intervention and population icons</td>
</tr>
<tr>
<td><img src="image" alt="Unidirectional Arrow" /></td>
<td>Unidirectional arrows for causal relationships</td>
</tr>
<tr>
<td><img src="image" alt="Bidirectional Arrow" /></td>
<td>Bidirectional arrows show feedback loops</td>
</tr>
</tbody>
</table>
## Quality of Execution Assessment Framework

<table>
<thead>
<tr>
<th>Domain</th>
<th>Potential Reasons for Limitations</th>
<th>Maximum Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>• Was the study population well described?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• Was the intervention well described? What was done? When it was done? How it was done? Where it was done? How was it targeted to the study population?</td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>• Was the sampling frame/universe adequately described?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• Were the inclusion and exclusion criteria clearly specified?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Was the unit of analysis the entire eligible population or a probability sample at the point of observation?</td>
<td></td>
</tr>
<tr>
<td>Measurement</td>
<td>• Were <strong>outcome</strong> measures valid and reliable?</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>• Was <strong>exposure</strong> to the intervention assessed?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• If yes, were these exposure measures valid and reliable?</td>
<td></td>
</tr>
</tbody>
</table>
## Quality of Execution Assessment Framework (cont’d)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Potential Reasons for Limitations</th>
<th>Maximum Limitations</th>
</tr>
</thead>
</table>
| **Data Analysis**             | • Appropriate statistical testing conducted?  
                                   • Reporting of analytic methods and tests?  
                                   • Appropriate controlling for design/outcome/population factors?  
                                   • Other issues with data analysis                                                                                                                                                | 1                   |
| **Interpretation of Results** | • >80% completion rate? Data set complete?  
                                   • Study groups comparable at baseline? If not, was confounding controlled before examination of intervention effectiveness?  
                                   • Biases that might influence the interpretation of results including other events/interventions that might have occurred at the same time.                                                                 | 3                   |
| **Other**                     | • Other biases or concerns not included in the previous domains (e.g., evidence of selective reporting)                                                                                                                             | 1                   |