Advancing Real-World Prevention with Statewide Data Systems: Insights from Washington State's Minerva 2.0

Heather Terral, PhD,
Gitanjali Shrestha, PhD,
Jaymie Vandagriff, PhD,
& Brittany Cooper, PhD

Washington State University & Washington Health Care Authority

May 30, 2025

Washington State Health Care Authority

IMPACT IMproving Prevention through ACTion Research Lab

WASHINGTON STATE UNIVERSITY

Agenda

- Describe the process of linking multiple large datasets containing real-world reporting data into a single analyzable dataset
- Report overall outcomes of that process
- Share lessons learned collaborating with state partners to utilize participant data collected for universal prevention

Welcome & Introduction

Who is in the room?

Snap if you’re familiar with relational databases

Clap if you’ve seen evaluation results from relational databases

Overview of Washington State's Approaches to Substance Use Disorder Prevention

Washington's Prevention Approaches

- Evidence-Based Program Delivery
  - **Community-Based Organizations (CBOs)**
  - **Community Prevention and Wellness Initiative (CPWI) coalitions**
  - CPWI School-based services

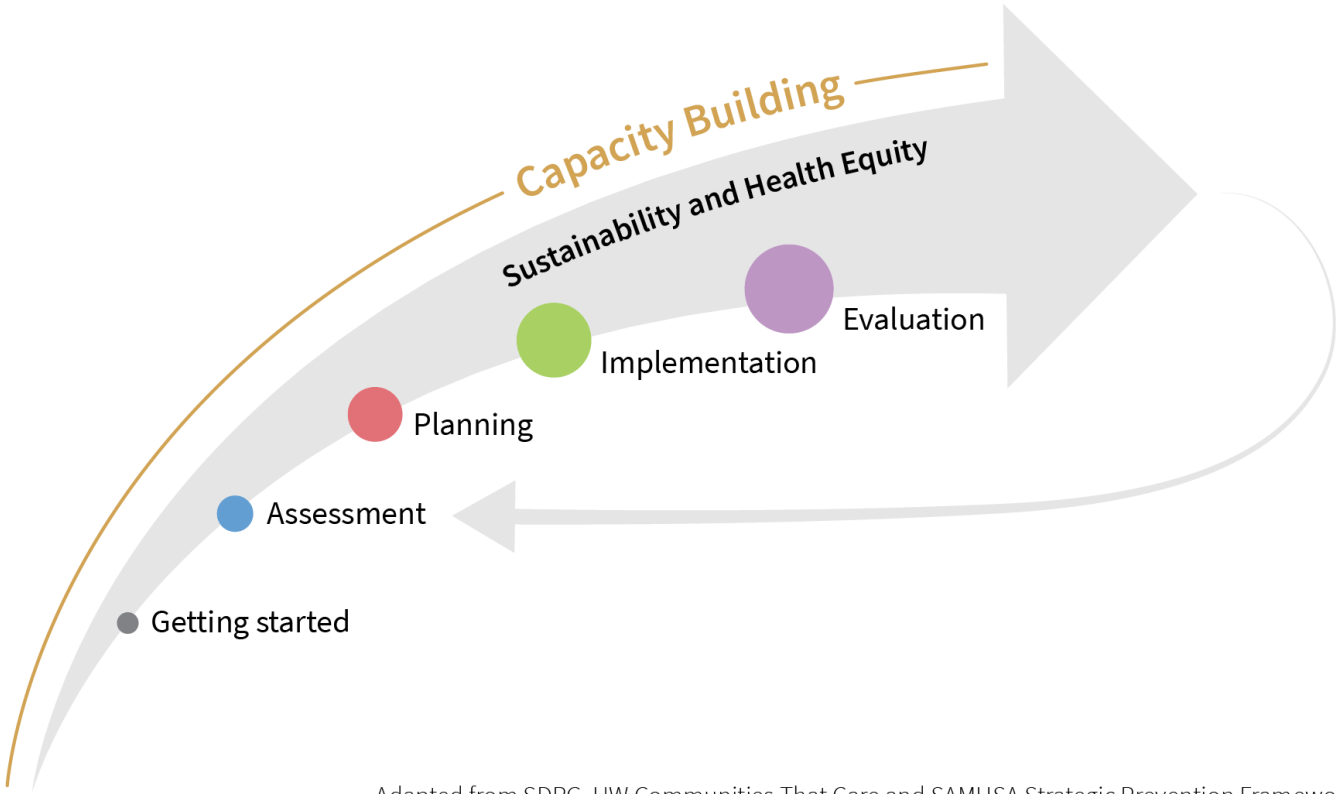
The CPWI Model

Washington State Health Care Authority

CPWI Strategic Planning Framework

- CPWI coalitions follow a strategic planning framework adapted from the SPF and Communities That Care (CTC)
- CBOs implement and evaluate EBPs
- Washington needed a reporting system that follows the SPF and can meet complex reporting needs

Washington State Prevention Planning Framework

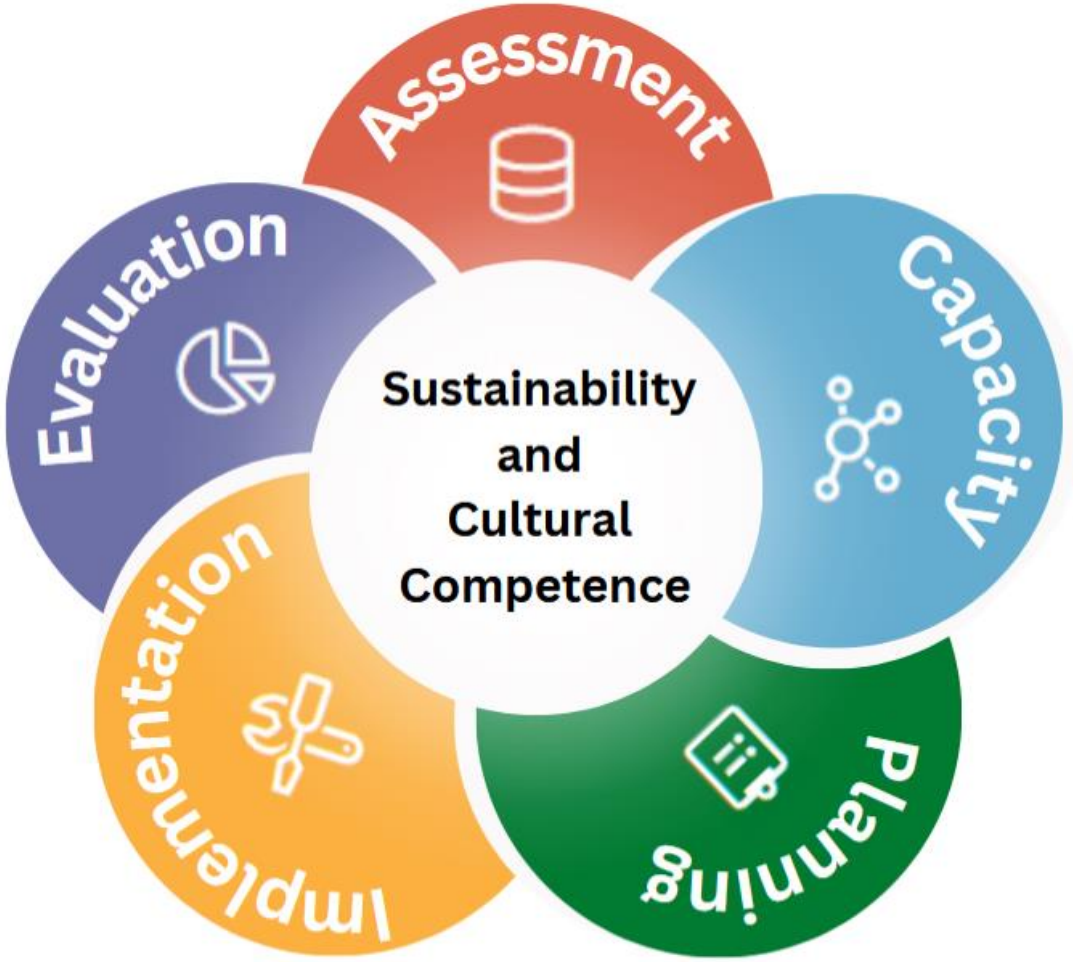


Adapted from SDRG, UW Communities That Care and SAMHSA Strategic Prevention Framework

Washington State Health Care Authority

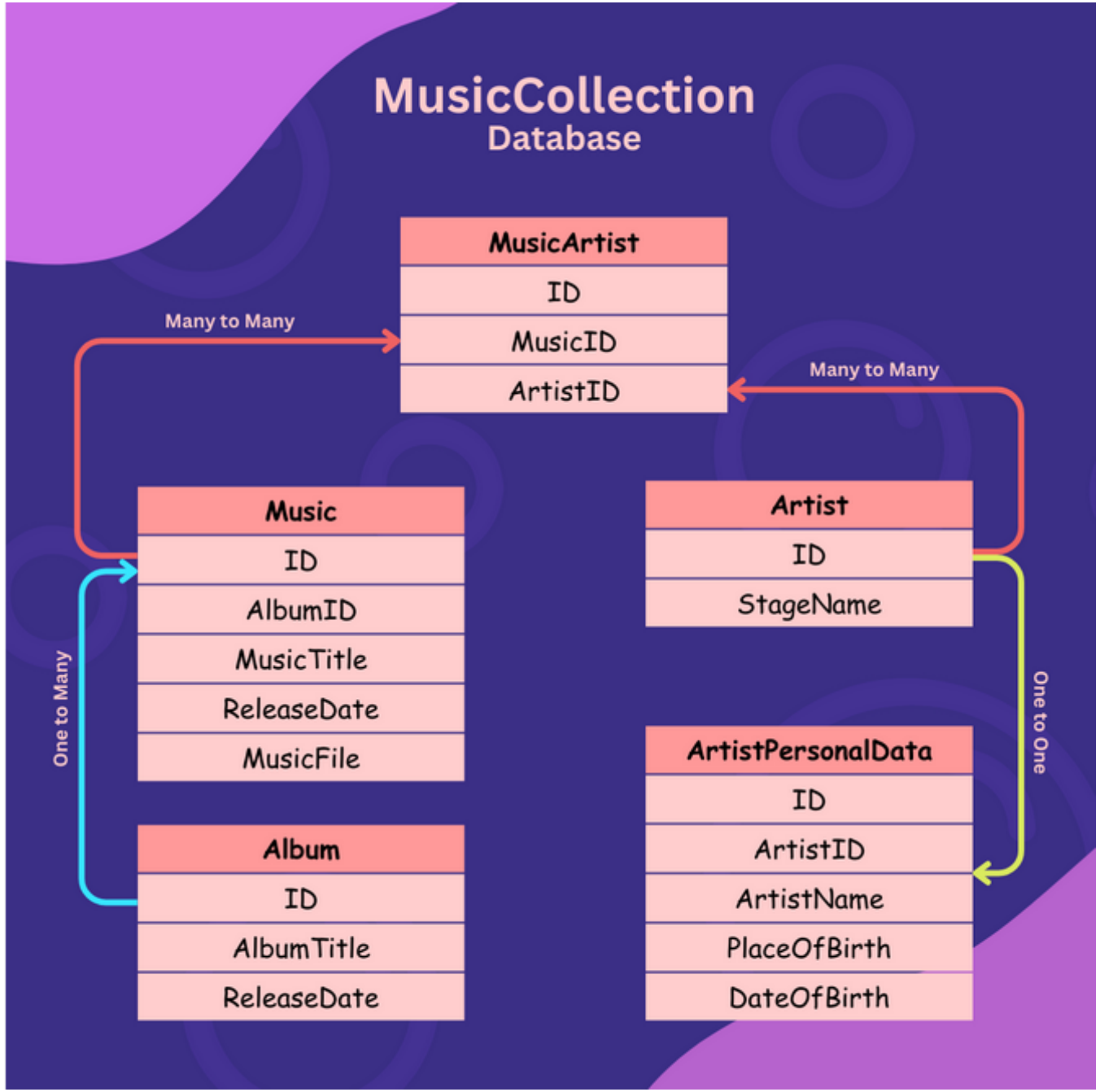
Meet Minerva 2.0, Washington's PBPS

- Prior reporting system was given the user-friendly name *Minerva*
- In 2022, Washington launched the Performance Based Prevention System (PBPS) database, *Minerva 2.0*
- Developed by Collaborative Planning Group Inc., the database structure reflects components of the SPF
- Each petal of the SPF corresponds to abilities within the database



(CPGSI, 2024)

Relational Databases



https://codeinpurple.com/example-of-a-relational-database/

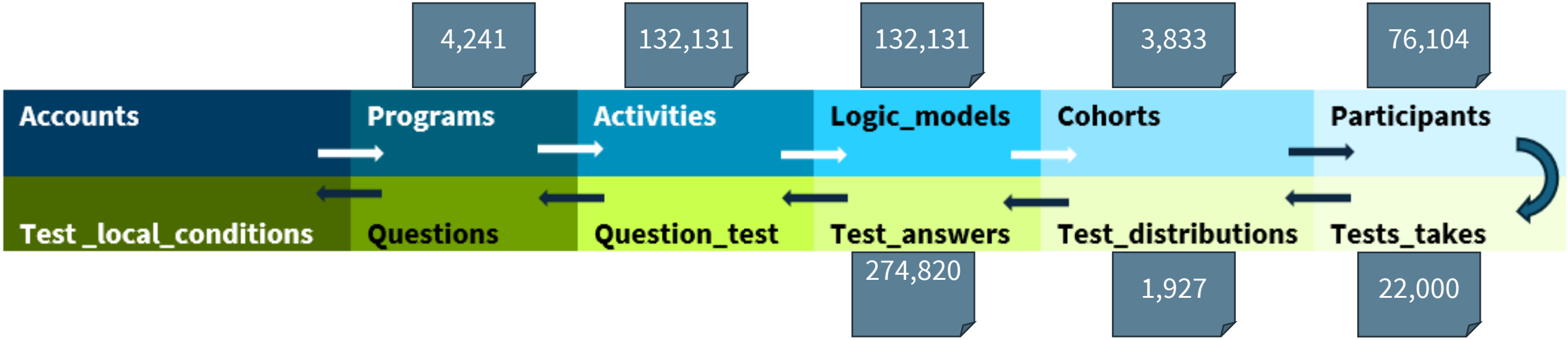
Methods

Project Objective and Process

- Began in October 2023
- Data points included program participation, fidelity, and prevention-related outcomes
- To get from 12 data tables to one dataset required various steps including linking, validation, matching, and patching
- RStudio was used to write the data wrangling syntax and then ultimately analyze those data

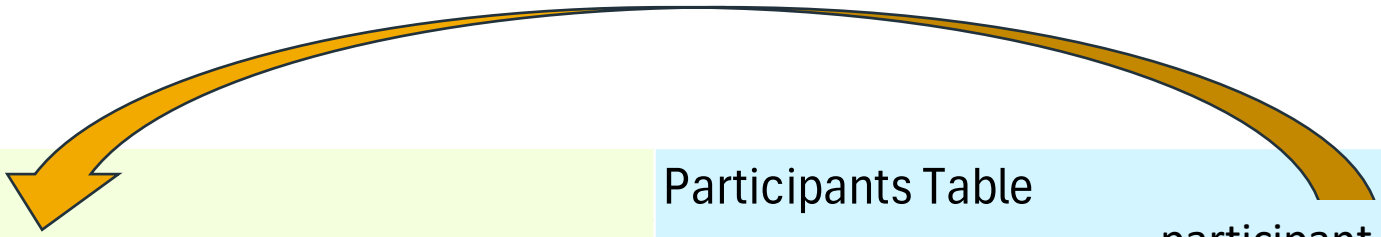
Project Objective and Process

- **Goal:** connect **participant** survey responses to the **programs** implemented, along with the **account** (entity responsible for contracting) that delivered the program
- Given the structure of the data, survey-related data had to be joined (questions, answers, test takers, unique test ids)
- Once that was accomplished, participant responses could be connected to the program that administered the surveys



Missingness

- Test_taker_ids are uniquely identifiers for participants
- Unfortunately, not all pre-post responses had test_taker_ids – 35,286 rows (33.2%) had missing data
- There were also many missing values for cohort_id
- This was a substantial portion of missingness and to recover those rows, a patching procedure was conducted



Test_Takes Table

first_name	last_name	test_taker_id	concat
Aang	Airbender		AangAirbender
Katara	Waterbender		KataraWaterbender
Toph	Earthbender		TophEarthbender

Participants Table

first_name	last_name	participant_id	concat
Aang	Airbender	123	AangAirbender
Katara	Waterbender	234	KataraWaterbender
Toph	Earthbender	345	TophEarthbender

Patching into the Analytic Dataset

- Needed to add matched rows into the analytic data set and check accuracy
- Of those responses that were matched by the concatenated name variable 3,094 (4%) were not exact matches between first and last names
- These were further examined to determine the nature of inconsistencies

Characteristics of inexact matches

- Visually inspected to confirm if data were appropriately matched in the patching process
- All participants names were the same between participants and test_takers, yet data entry error prevented exact matches
- Inputting program names or codes instead of participant names
- Inconsistent inputs for name or nickname (i.e, Abby or Abigail).
- Many of these names with discrepancies are not typically English-sounding names, which is an equity issue

Results

The Merged Dataset

- 44 columns x 73,470 rows, 3,232,680 total observations
- 73,470 survey item responses collected from 4,379 unique program participants across 280 pre-tests and 265 post-tests
- 121 program implementations
- Demographic items tended to be missing (optional fields)
  - Higher missingness rates for sexual orientation (87.6%) and transgender identity (100%) than for ethnicity and race (82.4%)

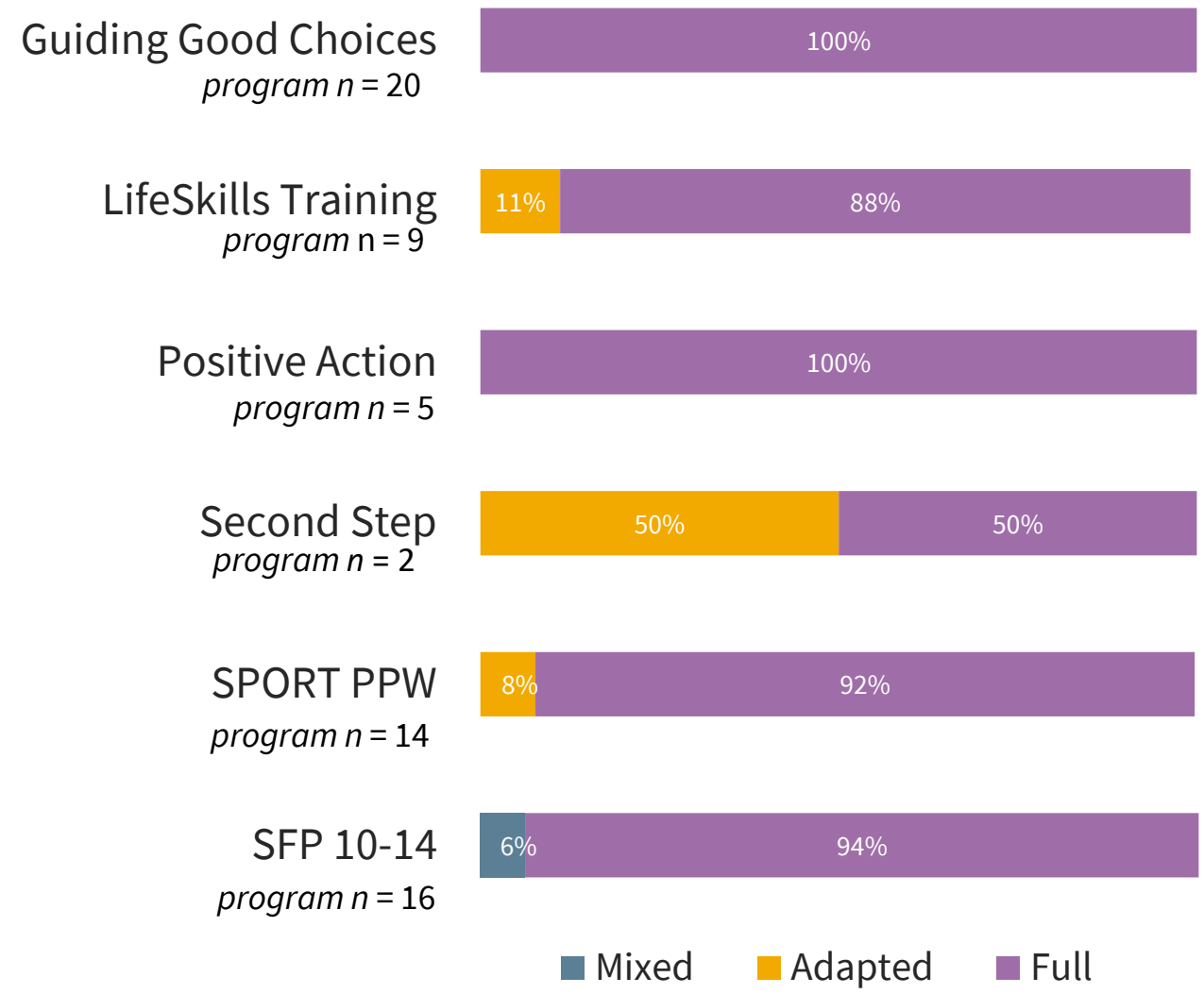
The Merged Dataset

- Was subset into 3 analytical datasets for evaluation
- Three reports produced:
  - **Adaptation/Fidelity** – Frequencies of EBPs being implemented with full fidelity
  - **CBO** – Participant outcomes of direct services provided by CBOs & their funding sources
  - **CPWI** - Participant outcomes of direct services provided by CPWI & cohort-level descriptives

Program Name	n
Guiding Good Choices	24
Strengthening Families and Youth 10-14	23
SPORT Prevention Plus Wellness	14
LifeSkills Training	9

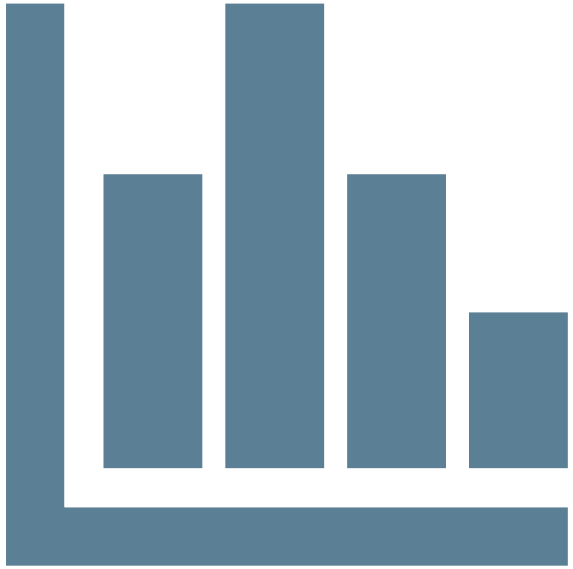
Adaptation/Fidelity Report

- Evaluation of most common EBPs
- Assessed extent of fidelity in implementations
- Most adaptations tended to be modifications to delivery (i.e., virtual implementation)



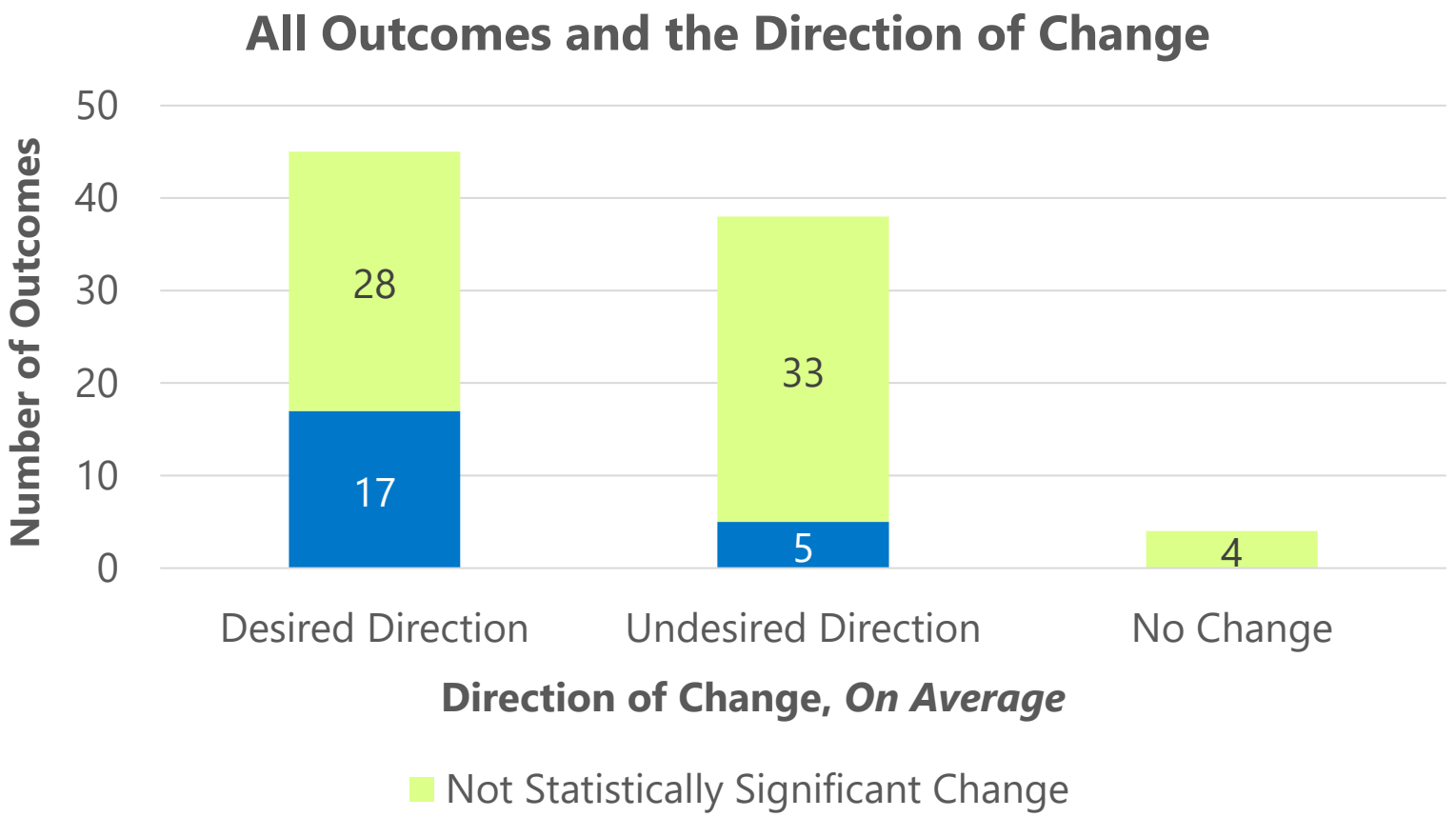
Community-Based Organizations Report

- Pre-post survey data for 7 programs implemented by CBOs (21 total implementations)
- 791 unique participants measuring outcomes across 22 scales
- Guiding Good Choices was especially high scoring
- Changes over time for Refusal Skills in the LifeSkills Training program was less than anticipated



CPWI Report

- 60,821 observations
- 2444 unique participants
- Overall, people participating in programs implemented in CPWI communities tend to improve on measured outcomes



Conclusions

Use of a comprehensive framework to support prevention at scale

- Despite challenges, ample data from diverse programs and communities across Washington provided insights into different aspects of prevention programming
- Administrators were able to assess implementation fidelity at the program level
- Program-level insights were obtained for many direct service programs

Lessons learned for statewide prevention data systems

- Consistency is key for harnessing the power of relational databases
- Emphasize the equity aspect of data entry
- Encourage those entering data to take extra care when entering data with names from outside their culture, with multiple and hyphenated names, initials, and generational suffixes (i.e., Jr, Sr, III)
- Consider positive reinforcement/internal recognition for outstanding data quality
- Consider user interface (UI) options to manage data error

Thank you!

Heather Terral

heather.terral@wsu.edu

Jaymie Vandagriff

Jaymie.vandagriff@hca.wa.gov

Gitanjali Shrestha

gshrestha@wsu.edu

Brittany Cooper

brittany.cooper@wsu.edu