Transcript

From Screening to Shared Decision-Making: Use of the SWYC in Clinical Care

0:01

I'm excited to report that as of 2025, the SWYC has a new home with the

TEAM UP Scaling and Sustainability Center. TEAM UP's goal is to transform care, empower families, and strengthen community by promoting the integration of comprehensive behavioral healthcare into primary care pediatric settings. I co-developed the SWYC with the same goals in mind, and I'm excited to tell you more about what it means for TEAM UP to be the SWYC's new home. I'll get to that--But first, a little background.

What is the SWYC, anyway?

And how does the SWYC help to achieve these goals?

0:45

First of all, SWYC stands for Survey of Wellbeing of Young Children. The SWYC is a set of screening questionnaires designed for children from birth to 5 years. Now the SWYC is not the only option--there are many screening questionnaires for kids in this age range that are well designed and have a strong evidence base. So what makes the SWYC different? When Dr. Ellen Perrin and I co-developed the SWYC at Tufts Medical Center, we had a set of design principles in mind.

1:18

By design, the SWYC is designed to feasible in primary care settings, so it is:

Brief—typically < 10 minutes to complete

Reliant on parent report—and thus be easily completed at home or in busy waiting rooms

Freely available—thus saving resources for other areas of care

Modular—we recommend using the whole thing, but you can choose the parts that you need for your practice

Amenable to different formats—including paper-and-pencil, telephone, or electronic forms

Longitudinal—forms coincide with standard pediatric visits, allowing wellbeing to be monitored over time. Moreover, the SWYC

Has a strong evidence base—ongoing research suggests that the SWYC has good reliability and validity, and that its accuracy is comparable to other prominent screening instruments.

Prioritizes sensitivity—thereby bringing as many problems as possible to the provider's attention. Therefore, instead of using the term "positive screens", we prefer the term "requires attention" or "needs review" to emphasize that the SWYC is not designed to stand on its own, but instead to inform shared decision-making

2:35

Let's get down to specifics. As I said, the SWYC is designed for children birth through 5 years of age. It includes several elements

2:41

The first is the developmental screening part of the SWYC. This includes the SWYC milestones, which is designed to detect developmental delays and is administered at all ages.

The SWYC also includes the parents observation of social interaction, or "Posey" for short, which is designed to screen for autism and is administered from 16 to 30 months.

3:05

The second part of the SWYC and is designed to screen for emotional and behavioral problems. For younger children the lower the age of 18 months, parents complete the baby pediatric symptom checklist or "B-P-S-C." Older children 18 months and above complete the preschool pediatric symptom checklist, or "P-P-S-C".

3:32

Finally there is a section of questions about family risk, including a parent depression, use of drugs and alcohol, family discord, and parent concerns. This set of questions is administered at all ages

3:48

Together these elements comprise the SWYC.

In practice we have different SWYC forms for every age on the American academy of pediatrics periodicity schedule from two months through five years of age. [click]

4:09

Here's an overview of the 18 month form as an example. As is true of all the 12 age-specific SWYC forms, this form is one page, front and (CLICK) back.

The first part of every form is the SWYC milestones. Note that the milestones section always starts with relatively easy questions and ends with hard ones. Thus, providers should not expect children to complete all milestones at any given age. The scoring algorithms adjusts for this.

The first part of every form is the SWYC milestones. Note that the milestones section always starts with relatively easy questions and ends with hard ones. Thus, providers should not expect children to complete all milestones at any given age. The scoring algorithms adjusts for this.

The second section of every form screens for emotional and behavioral problems. Because this is the 18 month form, the child receives the PPSC, which includes questions about anxiety, attention, and disruptive behaviors.

For children between 16 and 30 months of age, the next section is the POSI, which includes seven questions about autism risk

Next, there are two questions about parents' concerns that are recommended by the American academy of pediatrics

Finally there are nine questions about family risk. This includes questions about tobacco use, drug and alcohol use, food insecurity and hunger, PHQ 2 which is a well validated screener for parent depression, and the W-A-S-T, which screens for family discord and is sensitive to family violence. Note that the SWYC is designed to be modular. While we recommend using all parts of the SWYC at all ages, you can choose which parts best meet the needs of your particular pediatric practice. Also note that while I'm presenting an example of an English-language form, we are fortunate that colleagues have translated the SWYC into a number of other languages as well.

6:04

Ok, now that you know what the SWYC is, I'm ready to explain how I think the SWYC can help. If transforming care, empowering families, and strengthening communities seems like a lot to ask of a screening questionnaire, I'd agree: Dr. Perrin and I always said that screening is only one element in a larger system of care. So its important to understand

how we think the SWYC should be used—and how this differs from how screening is often implemented. To explain, let's take a step back and risk asking a silly question: what is a screening questionnaire?

6:54

Flash back for a moment to 1989 and 1990, which were watershed years for the science of screening. Based on an exhaustive review of the evidence and careful simulation modeling of its implications.

Ok, now that you know what the SWYC is, I'm ready to explain how I think the SWYC can help. If transforming care, empowering families, and strengthening communities seems like a lot to ask of a screening questionnaire, I'd agree: Dr. Perrin and I always said that screening is only one element in a larger system of care. So its important to understand how we think the SWYC should be used—and how this differs from how screening is often implemented. To explain, let's take a step back and risk asking a silly question: what is a screening questionnaire? Flash back for a moment to 1989 and 1990, which were watershed years for the science of screening. Based on an exhaustive review of the evidence and careful simulation modeling of its implications,

What was meant by screening? David Eddy's book,

"common screening tests," articulated the prevailing view in medicine, which defines Screening as

"the application of a test to detect a potential disease or condition in a person who has no known signs or symptoms of that condition at the time the test is done ." In this sense, screening works a lot like radar.

7:37

Just as radar is designed to detect flying objects like airplanes long before they can be seen or heard, screening is designed to detect a potential disease long before a person has any signs or symptoms. Screening for things like cancer or heart disease often work this way.

But both screening tests and radar sometimes make mistakes.

7:59

Sometimes the plane does not give off a signal, which is an example of a false negative error.

8:11

Other times, radar detects a signal, but it turns out to be something you're not really interested in detecting—like a flock of birds. This is an example of a false positive error.

8:16

With a sky full of flying objects, ideal radar will detect all of the planes but none of the birds. To evaluate the accuracy of radar, scientists developed something called

8:37

Signal detection theory, which is used to this day—not just for radar, but also for medical screening tests. For example, groups like the united states preventive services task force rely on signal detection theory when they carefully assesses the sensitivity and specificity of a screening test before recommending it.

8:48

But are developmental behavioral screening questionnaires really screeners in this sense? Sort of. On the one hand, medical professionals certainly

send out screening questionnaires to patients in the hope that they will answer the questions and return them with information about developmental and/or behavioral problems that their providers don't yet know about. But to answer the questions, families must be aware of the signs and symptoms, even if they don't recognize them as such. In this sense, the information they provide isn't really new – – at least not to the patient. It may be new to the provider, in which case, we have to ask, is the screening detecting signals, or is it more accurate to say that it is facilitating communication?

Radar doesn't seem to be the greatest metaphor—It's more like a conversation.

9:44

I imagine it more like this:

You called someone to ask them if they are on a plane

They don't have the expertise you do, so they ask for help

You suggest looking out the window, at which point

Their child exclaims that they are way up in the air. To complete this metaphor, imagine you have to do it again and again with a lot of people, so you start the conversation with

10:11

a standardized questionnaire. Now, this example may not reflect how we go about detecting planes these days, but it does seem a lot more like how medical professionals screen for developmental and behavioral problems than radar does

10:25

To summarize, there are at least two different ways of thinking about screening. One is like radar—the Signal detection way. If you buy into this way of thinking, you may assume that screening offers the only good source of information about mental health, and so you care a lot about whether or not screening questionnaires are "validated"—and in particular whether they accurately detect the problems you're looking to screen for. This has implications for how the screener is used.

10:55

Here's an example.

1.In 2006, the American academy of pediatrics published guidelines for developmental screening.

2.They said that after administering a validated screening tool,

3. Clinicians should determine whether the results are positive, and if so

4. Make referrals for appropriate care. While the figure you're seeing comes from a 2006 recommendation, this model is quite common in the field. For example, many implementation studies expect all positive screens to result in referrals, and the United States preventive services task force conceptualizes screening questionnaires this way. In short, this model suggests that decisions about what to do rest entirely on screening results.

11:41

But there's a second way to think about screening—

1. Screening As an evidence-informed conversation starter

11:48

For example, a newer screening recommendation from the American Academy of Pediatrics published in 2025 stated that screening and surveillance both present an opportunity to initiate conversations with children and families. Notably, this way of thinking suggests a much bigger role for both the clinician and the family. Its not as simple as whether or not the screen is positive.

12:14

And it also has implications for what research evidence is useful.

EBM has been defined as “the ability to track down, critically appraise…, and incorporate [emerging research] evidence into one's clinical practice”. In this regard, evidence about accuracy is certainly important—no doubt about it—but so is a range of other research evidence, such as normative data to interpret how extreme or unusual a score is, and data on how individual items work to help clinicians and families interpret results.

12:55

After all, evidence-based medicine isn’t just about

Evidence. According to the founders of the field, it is also

about applying evidence with clinical expertise

And its also about making clinical decisions in light of patients’ values and preferences,

These three concepts came to be described as a “three legged stool.” Just as a stool will fall over if any one leg brakes, clinical decision-making will fail if any of these concepts is neglected. Thus, evidence-based practice is not static—its a process,

…a process that is not "cookbook" medicine. External clinical evidence can inform, but can never replace, individual clinical expertise. To adapt a famous phrase, cookbook medicine runs the risk of prescribing a foolish consistency that is the hobgoblin of small minds. And so its important to think about screening as a process, a process that is but one element in a much larger system of care, and one that requires careful judgment and consideration of both evidence and values. Bear with me as I dig a little deeper to explain why.

13:56

To do so, I’m going to ask you to imagine a screening questionnaire that yields scores between 0 and 9. Note that this is a simple model—a thought experiment, if you will. While it doesn’t correspond exactly to any real-world screener, I think you will find the resemblances to be informative. I’ll point out some key concepts along the way.

14:19

As with many screening questionnaire, Parents answer questions, clinicians add up the scores, and those scores fall over a wide range

14:44

We refer to this range of scores a distribution.

What can we say based on this distribution? For starters, we can say something about how common or rare a score is, ideally using plain language.

15:04

For example, if a child receives a score of 5, we can say that the child scored in the top 27.9%. Better yet, we might say that out of 100 families, the parent reported more symptoms than about 72. Research supports plain language, and plain language facilitates conversation.

15:22

If a child receives a score of 6, we can say that out of 100 families, the parent reported more symptoms than about 91.

15:35

But we can say more.

It is often useful to think about kids as falling into two categories. Here in blue is the first category—those children who do not have developmental-behavioral problems, who I'll refer to as typically developing.

And here in red is the second category—those children who have developmental-behavioral problems. What can this tell us?

16:08

If the screener is supported by a good diagnostic accuracy study, we will know how many kids with a given score—in this case a score of 5—

Have a developmental-behavioral problem and

How many do not. Based on this evidence, we might say that:

"among children who scored 5 in a recent study, 27% were found to have developmental-behavioral problems"

"among children who scored 6, 62% were found to have developmental-behavioral problems"

17:40

And "among children who scored 7,

88% were found to have developmental-behavioral problems". This kind of information—i.e., probabilities associated with specific scores—is clinically useful.

So let's look at this same data in a different way that emphasizes positive and negative scores. So far in this example, children with problems were folded into the overall distribution…

But we can pull them out

as their own independent distribution…

And we can look at their scores alongside the distribution for typically developing children. **17:47**

We can now see that, typically developing children have an average score of 4,

whereas children with developmental behavioral problems have an average score of 5.5. On average, children with problems score higher—if they didn't, the screener wouldn't have any accuracy at all. But as you can see, there's a lot of overlap

You can see the overlap even better if we flip the red distribution below the horizontal axis. Now we're ready to consider the last element that makes this an example of real screening instrument—

the cut score, also known as a threshold, which determines which scores

indicate a positive screen

Versus a negative screen

Among kids with developmental-behavioral problems, this distinguishes

true positives – that is, kids with problems who we want to screen positive -- from

false negatives – that is, kids with problems who falsely screened negative. Likewise, among typically developing children, the threshold distinguishes between

True negatives – that is, kids without problems who should have screened negative – from

False positives – that is, kids who screened positive even though they had no problems

Here, I just shade the correct results—true positives and true negatives. Notice that a lot of the scores aren't shaded in—there are plenty of incorrect screening results here. Let's ask a few questions to see how accurate this screener is

Consider the children with problems who fall below the horizontal axis. What percentage are correctly classified as having a problem? That is, what proportion of scores fall in the "positive" range to the right of the threshold? Look at the figure and think about this for a moment. [click]

If you said “two thirds” or “three quarters”, you are very close. The actual number is 78%.

This is what we mean by “sensitivity”. Sensitivity is the percentage of children with problems who are correctly classified by the screener.

20:00

Now consider the typically developing children. What proportion are correctly classified as not having a problem? That is, what proportion fall to the left of the threshold in the negative range? Look at the figure and think about this for a moment.

Again, If you said “two thirds” or “three quarters”, you are very close. The actual number is 78%.

20:36

This is what we mean by “specificity”. That is, specificity is the percentage of children with NO problems who are correctly classified by the screener. This figure depicts scores on a screener that displays 78% sensitivity and 78% specificity. Most people would say this is a good screener. But is it good enough to make clinical decisions? Let’s take a closer look.

20:50

To think about this screening instrument from the perspective of a clinician, we might ask a different question. When a clinician cares for a child with a positive screen, the clinician would ideally like to know the chance that this child has an actual problem. [click]

So consider the children who score positive. These are the scores that fall to the right of the threshold in the positive of range, including those above and below the horizontal axis. Which positive scores are correct? That is, which scores co-occur with an actual problem? These are the scores below the axis. Among all children with positive scores, what percentage fall below the axis? Look at the figure and think about this for a moment [click]

If you said about half, then you are very close. In fact, less than 50% of children who score positive have actual problems. [click]

This is what we mean by positive predictive value, or PPV. So, if we implemented a policy that all children who score positive on this screener should be referred, then we would be wrong more than ½ the time, even though we are using a good screener to detect a highly prevalent condition.

22:05

But the situation is actually more complicated. We know that children with higher positive scores have higher risk—actually risk that is much higher than the PPV, whereas others have moderate risk

And still other children—namely, those who score just above the threshold—are at much lower risk than the PPV would suggest.

The issue is that PPV averages risk across all who score positive. Yet we know that risks associated with individual scores vary widely. PPV may be useful for understanding groups of children with positive scores, but in the end, clinicians and parents tend to care about one child at a time. And each of those children has only one score at a time, and one particular prognosis.

Now remember, this is only a model. We should proceed with caution. That said, it is useful in at least two ways.

In short, this model tells us something about validated screening instruments. As I mentioned, sensitivity and specificity

Are 78% in this example. So this is precisely the kind of screener that is typically described as "validated." And yet,

22% of children are misclassified. And if we believe the claim that Pediatricians should rely on validated screeners like this one when making referral decisions, not their clinical judgment, then we need to accept that

More than 50% of referrals will be false positives, with an even higher risk for children who score positive right at the threshold. Are validated screening instruments really sufficient by themselves to drive clinical decisions in this way? I think not, and that's

Key concept #4.

Now you may be thinking, ok fine, not all validated screening questionnaires can justify referrals, but the better ones can. Point taken. Maybe some screeners are highly accurate, like this:

And I'd say sure—that's possible, and if you can find me a screening questionnaire as accurate as that, I'd love to see it. But what I see more often is screeners like the one I've been describing,

But without mention of predictive value

Without mention of errors at all

Emphasizing sensitivity and specificity, sometimes replacing it with some kind of

odds ratio between positive and negative screens (which is 11.4; often I see claims for screeners with odds ratios of 4 and under)

And certainly not including a visual of the distributions. As a result, I think that recommendations to use so called validated screening questionnaires as if they were radar—often based on limited information--is just another modern just so story. We can do better. For example, we can be clear about the limitations of even the most valid screening questionnaires, and we can strive to offer clinicians and families all the information they need to make informed decisions.

There's a second way in which this screening model is useful: it highlights the role of values in healthcare.

After all, there tradeoffs are inherent in any clinical decision—not just for developmental behavioral screeners – but for any medical test or any decision for that matter. As we raise the threshold,

26:03

Fewer children screen positive. This means that fewer children with problems screen positive, which reduces sensitivity. That's bad. But there are tradeoffs. It also means that fewer typically developing children screen positive, which raises specificity. That's good. It also means that more—or in this case most—of the kids who screen positive actually have problems. That is, PPV is high. That is also good. We see different tradeoffs when we

Lower the thresholds. More kids screen positive. This means that more children with problems screen positive, which increases sensitivity. That's good! But again there are tradeoffs. It also means that more typically developing children screen positive, which reduces specificity. That's bad. It also means that fewer of the kids who screen positive actually have problems. That is, PPV is low. That is also bad. As I said, thresholds are a general feature of many decisions, not just medical tests and not just developmental behavioral screening. They also vary among different tests. Some screeners have very high thresholds—this may seem good at first because a high proportion of the children who screen positive turn out to have problems. But remember there are tradeoffs—a high threshold typically implies lower sensitivity, which means the screener is missing kids you would like to detect. Conversely, some screeners—like the SWYC-- have lower thresholds. This may seem bad at first because they can yield a lot of false positives—but remember that the threshold was set where it was to ensure that sensitivity was comparatively high.

Given all of these tradeoffs, how do we know where the threshold should be? A lot of researchers and most screeners rely on something called Youden's index, which frankly

makes no sense for reasons that go well beyond the scope of this talk. Suffice it to say that its overly technical and ignores the importance of values.

28:13

Another option is something called decision analysis, which embraces the importance of values. Alas, it is perhaps even more technical, so allow me to summarize the main points.

What decision analysis says is that if it truly makes sense to refer a child who scores above the threshold, this implies that the benefits of doing so

Outweigh the costs. In contrast, if it truly makes sense for a child who scores below the threshold to NOT be referred. This implies that the costs outweigh the benefits

At some point in between, costs and benefits are in balance. This is the point of indifference, and by the logic of decision analysis, this is where the threshold should be. Let's consider two examples.

29:38

Let's start with cardiology. The USPSTF suggests that for those at risk of a cardiovascular event,

The benefits of taking statins far outweigh the risks and costs. So,

Where should the threshold be? Well, given their estimate of costs and benefits, the USPSTF says that the threshold should be

Quite low—even as low as

8-10%. That is, if an individual has even an 8% risk of a cardiovascular event in the next 10 years, they should consider taking statins

29:53

Let's start with cardiology. The USPSTF suggests that for those at risk of a cardiovascular event,

The benefits of taking statins far outweigh the risks and costs. So,

Where should the threshold be? Well, given their estimate of costs and benefits, the USPSTF says that the threshold should be

Quite low—even as low as 8-10%. That is, if an individual has even an 8% risk of a cardiovascular event in the next 10 years, they should consider taking statins. Now you may agree or disagree—the USPSTF is quite clear that while these estimates are based on

a host of evidence, they are also the product of expert judgment. And there is certainly room for shared decision-making with patients. But for now, let's contrast this with a different case: where should we set the threshold for an expensive and invasive surgery?

Even if the potential benefits are the same,

the risks are considerably higher,

As are the costs

So the balance between the two is different. Does surgery make sense if the risk of a cardiovascular event is only 8-10%? Given the risks and costs, probably not

30:58

A much higher threshold probably makes sense. So what have we learned by this example? Notice that I said nothing about how we are screening for cardiovascular risk—presumably we're using the same risk calculator in both examples. What changed is the treatment being considered, and more specifically, the risks, costs and benefits that are likely given that treatment. So an optimal threshold isn't really a property of a screening tool—it has much more to do with the net benefit of available options. That is, it depends on the context in which the screening occurs. In my opinion, this is

Another Key concept. Optimal thresholds require weighing risks and costs with benefits. And while this process benefits from evidence, it clearly requires judgment and values. We can think of other examples that are specifically relevant to pediatrics. One such example is screening for risk of early onset sepsis among newborns. The same logic applies, and the recommended threshold for administering antibiotics is again very low.

32:13

So how can we extend this same logic to clinical decision making for developmental behavioral problems in primary care pediatrics? Screening can help us estimate symptoms and risk of disorder, but what severity of symptoms or what probability of disorder justifies further assessment or treatment? Decision analysis suggests that for each available option, we should consider all of the potential costs and all of the potential benefits. And in your system or for your family, who knows what options are available? Who has knowledge of the benefits? Who has a perspective on the costs? I'll tell you, it's not me. I can create screening tools that offer valuable information, but they can never replace the situated knowledge of families and their providers. Screening can never replace shared decision making. That's why we say that screening is only one element in a larger system of care.

And by recognizing this, we hope that the SWYC can make its own contribution to transforming care, empowering families, and strengthening communities. And that's why I'm so happy that the SWYC has found a home with TEAM UP, which helps primary care pediatric practices develop care teams that include nurses and pediatricians but also licensed behavioral health clinicians and community health workers. From my perspective, TEAM UP offers exactly what families need to engage in meaningful shared decision-making, and so it is the perfect home for the SWYC.

33:23

TEAM UP has only hosted the SWYC for a short time, but already its made a big difference. You can now access the forms through the website, including translations. There's also a link to contact us with questions or requests for free licenses. And we have much more planned. Soon, the website will include a searchable list of research articles about the SWYC, many by independent investigators from other teams. In addition, we're planning a series of free videos that will go deeper into the concepts I've discussed today, and more. For example, we plan to have videos that take a deeper dive into how to use developmental milestones in clinical care and in research, as well as to our approach to translations. Thanks to TEAM UP, all of these resources will remain

Free. This is a big deal. Frankly, there's really no such thing as a free screening tool. SOMEONE has to maintain the website and answer the emails. And so the TEAM UP center now hosts the SWYC because the missions align. It's about transforming care, empowering families, and strengthening communities. I hope that the SWYC and TEAM UP can be helpful

THANK you! And please contact us if we can be of further help