The Potential for Improving the Population Health Effectiveness of Screening: A Simulation Study

William Gardner, PhD, Katherine Bevans, PhD, Kelly J. Kelleher, MD, MPH

abstract

BACKGROUND AND OBJECTIVES: Screening interventions in pediatric primary care often have limited effects on patients' health. Using simulation, we examined what conditions must hold for screening to improve population health outcomes, using screening for depression in adolescence as an example.

METHODS: Through simulation, we varied parameters describing the working recognition and treatment of depression in primary care. The outcome measure was the effect of universal screening on adolescent population mental health, expressed as a percentage of the maximum possible effect. Through simulations, we randomly selected parameter values from the ranges of possible values identified from studies of care delivery in real-world pediatric settings.

RESULTS: We examined the comparative effectiveness of universal screening over assessment as usual in 10,000 simulations. Screening achieved a median of 4.2% of the possible improvement in population mental health (average: 4.8%). Screening had more impact on population health with a higher sensitivity of the screen, lower false-positive rate, higher percentage screened, and higher probability of treatment, given the recognition of depression. However, even at the best levels of each of these parameters, screening usually achieved <10% of the possible effect.

CONCLUSIONS: The many points at which the mental health care delivery process breaks down limit the population health effects of universal screening in primary care. Screening should be evaluated in the context of a realistic model of health care system functioning. We need to identify health care system structures and processes that strengthen the population effectiveness of screening or consider alternate solutions outside of primary care.

WHAT'S KNOWN ON THIS SUBJECT: Many adolescent behavioral and mental health problems are underrecognized and treated in primary care settings, prompting proposals for universal screening. However, universal screening's population health effects are not well-known nor are the determinants of screening effectiveness well-understood.

WHAT THIS STUDY ADDS: Our simulations suggest that the population health effects of universal screening in primary care may be low because care processes are subject to failure at many steps. Therefore, failures to complete care process steps accumulate to degrade system performance.

We define universal screening in primary care as a policy in which policymakers intend to screen every patient who comes through the office door. Universal screening programs are often recommended when there is a disorder with an effective treatment but many patients with the disorder are not identified and go untreated. Many writers have argued that adolescent depression is such a condition.\(^1\)\(^-\)\(^3\) The argument for universal screening for adolescent depression is as follows. (1) Depression is a prevalent disorder in this age group, with many adverse long-term psychosocial outcomes.\(^4\) (2) It is underrecognized and, therefore, undertreated. (3) The primary care system sees many adolescents, is trusted by families, and focuses on early intervention. (4) There are (moderately) accurate screening tools; and, finally, (5) there are (moderately) effective treatments for depression. We believe that assertions (1) through (5) are mostly correct. However, do they reveal that universal primary care screening for depression will substantially impact adolescent population mental health?

Screening interventions are cost-effective in some pediatric settings, such as screening newborns for metabolic disorders.\(^5\) However, it is difficult to find examples in which patients have been shown to benefit from universal preventive screening in primary care practice. In adult primary care, in randomized controlled trials of prostate-specific antigen screening, researchers indicate that the effects of prostate-specific antigen screening on mortality, if they exist, are too small to be reliably detected, even in trials involving several 100 000 patients.\(^6\)\(^-\)\(^8\) What accounts for these disappointing screening trial outcomes? It is routinely acknowledged that, to improve health outcomes, screening must be followed by an accurate diagnosis and available treatment of disorders.\(^9\) In our view, in recommendations for screening and monitoring, it is too often assumed that treatments for conditions identified during screening are acceptable, accessible to patients, and reliably delivered\(^10\); and, in those implementing screening programs, it is rarely assessed whether these conditions hold.

In this study, we considered universal screening for adolescent depression in primary care, and, using a simulation, we examined how multiple health care delivery processes influenced the effects of screening on population health. To that end, we conducted a simulation study that explored this question: how well does a primary care system need to function for universal screening interventions to affect population health meaningfully?

**METHODS**

**A Process Model for the Identification and Treatment of Depression in Primary Care**

We built our simulation on a care delivery process model (Fig 1; see also Supplemental Information) in which we describe, in a highly simplified way, how a health care system delivers depression care in a primary care setting to an adolescent patient. In Fig 1A, we present the model for primary care practice without universal depression screening. We call this “standard assessment.” The critical distinction is that, under standard assessment, the practice does not intend to screen every patient, although screens may be used ad hoc. Under universal screening, the policy is to screen every patient, although some patients may not get screened in practice.

The standard assessment process begins when a patient visits the primary care practice (or does not). It continues with the clinician’s recognition (or not) that the patient is depressed (or not depressed). It ends with treatment (or not), in which “treatment” represents any evidence-based treatment of adolescent depression (eg, cognitive behavior therapy or antidepressant medication) and is more substantive than the brief counseling offered by primary care clinicians. Treatment, if it happens, usually happens outside that office (eg, through the youth attending an appointment with a psychotherapist or filling a prescription in the community). This means that the care delivery process is not just a primary care process; it involves the health system in which the primary care office is embedded. Why would treatment not occur, although the patient has been recognized as depressed? Patients or families may disagree with the clinician’s judgment. Alternatively, the clinician may be reluctant to raise the issue either because they anticipate such disagreement or believe that a more pressing medical need takes priority. Perhaps a clinician may recognize depression and decide that the best course is watchful waiting, but then there is no follow-up visit to assess the patient’s current state. Likewise, there can be problems conducting an assessment and acting on it in real-time during a brief visit. For example, suppose that the clinician asks the patient to complete a depression inventory to confirm the diagnosis. It can be challenging to get this completed and scored before the family has to leave and the next patient arrives.

When universal screening is implemented (Panel B), 2 routes lead to treatment. If a youth visits the practice, they are screened (or not). If they are screened, then
depression may be recognized if the screen returns a positive result, and then, the patient is treated (or not). However, the practice might not administer the screen. If so, there is a route to treatment, beginning with clinician recognition of depression and leading to treatment. Again, the patient must traverse at least 1 of these routes to receive treatment.

**Population Mental Health Effects of Universal Screening**

To assess the population health effects of universal screening, we contrast 2 simulated worlds: 1 in which the health care system practices standard assessment and the other in which the health care system practices universal screening. In each world, we observe a cohort of youth for, let us suppose, a year. In that year, a youth in a cohort has a visit or not, and, if they do, that visit will result in a process that results in treatment (or not). Thus, in each world at the end of the year, there will be 4 groups of youth: (1) not-treated youth without depression, (2) treated youth without depression, (3) not-treated youth with depression, and (4) treated youth with depression. We assign each youth an outcome reflecting the effect of the health care system on their mental health. For youth with or without depression who were not treated, that score is 0, meaning that, on average, the system did not influence their mental health. We assign an outcome of 1 to treated youth with depression. This outcome represents the effect size of depression treatment, averaged across the system’s mix of treatment modalities and rescaled to 1 for
convenience. We assumed that treated youth without depression experienced a small iatrogenic harm, denoted as $\omega$. These effects are likely small because a youth without depression referred to a specialist will, most likely, be recognized as healthy and not receive medication. Even so, this unnecessary mental health intervention will waste the time of the primary care physician making the referral, mental health specialist making the assessment, and family and patient, including time that could be used to address some other health concern of the patient. As such, this false-positive error has caused the patient a small harm.

Then, for either simulated world, the expected change for the year in a youth’s mental health is the sum of these outcome scores, weighted by the probabilities that the youth is in each of the 4 groups. Finally, the population mental health effect of screening is the expected change in mental health under universal screening minus the expected change under standard assessment.

To make this outcome easier to understand, we rescaled the population mental health effect to a percentage of the maximum possible difference between universal screening and standard assessment, denoted as $\Psi$. The maximum benefit from screening ($\Psi = 100\%$) occurs when every youth has a primary care visit, all the patients without depression are recognized and treated, and none of the youth without depression are treated. $\Psi = 100\%$ does not imply that all youth would be healthy; it just means that every youth with depression and no youth without depression received treatment. Some treated youth with depression will remain unhealthy, but the system has done what it can. $\Psi = 0\%$ means that universal screening is no better than standard assessment.

### The Simulation

For simplicity, we assumed that each youth in the population had, at most, 1 visit during the year. A more realistic simulation would include subsets of youth with multiple visits. Multiple visits would change the distributions of outcomes by giving youth more opportunities for treatment and exposures to iatrogenic harm, but, because only a subset of youth would be affected, including multiple visits would have only modest effects on the results.

The standard assessment and universal screening models are defined by a set of parameters. These include the prevalence of depression in the population. Prevalence estimates for adolescent depression vary greatly, from $<3\%$ in a study of cumulative diagnoses in Denmark$^{11}$ to 3.6% on the basis of results from the 2016 National Survey of Children’s Health,$^{12}$ to 7.5% 12-month prevalence in the National Comorbidity Study.$^{13}$ Our principal simulations use a prevalence of 4%. The Supplemental Information reveals that the results are robust, over a range of 3% to 10%.

In our primary analyses, we also assume that the iatrogenic harm (that is, the side effects) suffered by treated youth without depression is 0.01, that is, one-hundredth the magnitude of the benefit received by a treated youth with depression, a small value consistent with a recent evidence review.$^{14}$ In the Supplemental Information, we present the effects of a range of effects $0 \leq \omega \leq 0.1$.

In Table 1, we present the parameters that describe the probabilities that a patient will progress from one step to the next in the primary care processes defined in Fig 1. Because there is uncertainty about these parameters' correct values, in our simulations, we randomly sample values in a range defined by the lowest and highest values found in the literature. In Table 1, we provide these ranges and their sources in the literature.

Each instance of the simulation involved a random draw from these ranges. Then, we calculated the expected proportions of youth in each of the 4 groups for the universal screening and standard assessment worlds. From those groups, we calculated the score for the effect of universal screening. In summary, we simulated the effect of universal screening. In our primary analyses, we also assume that the iatrogenic harm suffered by treated youth without depression is 0.01, that is, one-hundredth the magnitude of the benefit received by a treated youth with depression, a small value consistent with a recent evidence review.$^{14}$ In the Supplemental Information, we present the effects of a range of effects $0 \leq \omega \leq 0.1$.

In Table 1, we present the parameters that describe the probabilities that a patient will progress from one step to the next in the primary care processes defined in Fig 1. Because there is uncertainty about these parameters' correct values, in our simulations, we randomly sample values in a range defined by the lowest and highest values found in the literature. In Table 1, we provide these ranges and their sources in the literature.

### RESULTS

We ran 10,000 instances of the simulation. In each simulation, we randomly selected each process parameter value from the ranges in Table 1. Then, we calculated for that simulation.

In Fig 2, we illustrate how youth with depression flow through the system under universal screening, determining the proportions of treated and untreated youth with depression. We set the transition probabilities to the midpoints of the sampling ranges defined in Table 1. Thus, 44.0% of youth had a visit. Of those with a visit, 49.5% were screened (ie, 21.8% of youth with depression in the population). The screen had 73.5% sensitivity, so 16.0% of the youth with depression in the population had a positive screen result. However, of the 50.5% of youth with depression who were not screened, the primary care clinician recognized depression in 28.0% (6.2% of the population). In
TABLE 1 Care Process Parameters Governing Simulation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Visit = y)</td>
<td>The probability that a youth will visit primary care.</td>
<td>0.33</td>
<td>0.55</td>
<td>Rand and Goldstein⁵⁰ found that 50% of US adolescents had a primary care visit in 12 mo, but only 33% had a nonacute care visit, in which most preventive care occurs.</td>
</tr>
<tr>
<td>P(R = y</td>
<td>D = y)</td>
<td>The probability of recognizing depression in a standard assessment, given that patient is depressed.</td>
<td>0.21</td>
<td>0.35</td>
</tr>
<tr>
<td>P(R = η</td>
<td>D = n)</td>
<td>The probability of not recognizing depression in a standard assessment, given that patient is not depressed.</td>
<td>0.84</td>
<td>0.91</td>
</tr>
<tr>
<td>P(T = y</td>
<td>R = y)</td>
<td>The probability that treatment will occur given that depression was recognized.</td>
<td>0.17</td>
<td>0.78</td>
</tr>
<tr>
<td>P(Scr = y)</td>
<td>The probability that the screen will be administered.</td>
<td>0.25</td>
<td>0.74</td>
<td>Chisolm et al⁵⁶ reported that in a large trial of universal primary care screening for mental and behavioral health problems, only 25% of patients seen were screened. Savageau et al⁵⁶ studied the court-mandated mental health screening in Massachusetts and reported screening in 74% of cases.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>The probability that a youth screen result was positive, given that he or she is depressed.</td>
<td>0.57</td>
<td>0.9</td>
<td>In a meta-analysis,⁵⁷ researchers found that the sensitivity of the PHQ-2⁵⁸ was 57%. The sensitivity of the PHQ-A is 73%.⁵⁹ Frühers⁶⁰ 22-question Children's Depression Screener has a sensitivity of 90%.</td>
</tr>
<tr>
<td>Specificity</td>
<td>The probability that a youth screen result was negative, given that he or she is not depressed.</td>
<td>0.76</td>
<td>0.9</td>
<td>In a meta-analysis,⁵⁷ researchers found that the sensitivity of the PHQ-2 was 76%. Frühers⁶⁰ 22-question Children's Depression Screener has a specificity of 90%.</td>
</tr>
</tbody>
</table>

PHQ-A, Patient Health Questionnaire Modified for Adolescents; PHQ-2, Patient Health Questionnaire 2.

total, then, 22.2% of youth with depression were recognized through one path or the other. But, of those, only 47.5% received treatment (10.6% of the population).

The net effect of losses at multiple steps in the care delivery process was that 89.4% of youth with depression in the population were untreated. Also, 3.1% of youth without depression in the population received treatment. Therefore, among 1000 youth in the population, a 4% prevalence meant that there were 40 who were depressed, among whom 4.2 were treated. Among the 960 without depression, 29.5 were treated.

Figure 3 is a histogram of the 10 000 Ψ values generated in the simulations defined in Table 1. The median Ψ was 4.2%, and the average was 4.8%. A total of 61.5% of Ψ values were <5%, and the maximum Ψ was 19.5%.

In Fig 4, we present 4 scatter plots from the simulations, showing the effects of the primary care process’s quality (on the horizontal axes) on Ψ (the vertical axes). Figure 4 reveals that increasing the sensitivity of the screen (Fig 4A), reducing the false-positive rate (FPR) (Fig 4B), increasing the percentage screened (Fig 4C), and raising the probability of treatment will all increase the effectiveness of universal screening. What is surprising, however, is how small these effects are. Even at the best levels of each of these parameters (eg, sensitivity = 0.90), most of the values of Ψ are <10%. The consistently low values occur because one cannot accomplish much by fixing just 1 problem in a chainlike process that is fallible at each step. The probability of
delivering the service is always less than the lowest transition probability, and one has to get each step right to succeed.

**DISCUSSION**

We ran 10,000 simulations of the effect on population mental health of implementing universal primary care screening in a health system while holding other health care delivery parameters constant. Routine screening produced only small effects on population mental health: on average, <5% of the maximum achievable effect. Universal screening had greater effects on population health when the care delivery process had higher screen sensitivities, lower FPRs, higher percentages screened, and higher probabilities of treatment, given the recognition of depression. However, even at the best levels of these parameters, most simulated universal screening programs achieved <10% of the maximum possible population health effect.

Why does universal screening in primary care have such a modest impact on population health? The reason is that success is needed at each step of the process to deliver treatment. Therefore, all process factors combine multiplicatively, and the value of the smallest transition probability will be the upper bound on performance. Substantial losses at 1 step or modest losses at several steps produce poor overall performance. Empirical studies of primary care and the mental health care system are replete with evidence that such losses are frequent.

**Implications for Screening**

It follows that the implementation of universal screening in primary care should be preceded by a careful assessment of the system’s ability to deliver care. In many systems, universal screening must be accompanied by improvements in the health care system to succeed. First, we must address inadequate access to well-care among US adolescents, especially for minorities. One-half of adolescents do not have primary care visits in a year, only one-third have a well visit to primary care, and those who lack resources (eg, insurance) may be more vulnerable to depression. Quality improvements in primary care practices are essential so that they can reliably administer screens. Screening and assessment procedures are needed to improve the sensitivity and specificity of primary care diagnoses. Finally, it is essential to solve mental health care delivery problems. Many communities lack access to specialty mental health or substance abuse services. Primary care practices struggle to maintain ongoing patient registries for longitudinal tracking of referrals and laboratory data, motivating patients to engage in mental health care, and facilitating referrals to specialty care. Some positive outcomes of primary care-based depression screening have

![Depressed patient flow under universal screening](image-url)
been observed in adults but only when practices have additional support services. These services include facilitated referrals, patient tracking and monitoring, and treatment evaluation.20,21

Such efforts will pose significant challenges. Primary care practices are under increasing pressure to improve the delivery of chronic care services, address the social determinants of health, engage communities with digital tools, increase access to care, and conduct screening and other preventive activities for their patients, all while reducing costs.22,23 It is easier to set these mandates than fulfill them. In many trials and observational studies, researchers have documented the shortcomings of primary care practices in administering universal services like screening and associated treatments.24 In a study of practice improvement for attention-deficit/hyperactivity disorder, pediatric primary care groups increased their initial screening and prescribing rates, but long-term tracking, monitoring of schools, and treatment titration required significant support from our research team.25

Unfortunately, most individual primary care practices cannot implement effective universal screening and ensure the subsequent treatment of depression without extensive external support. That support must include the health care system’s ability to deliver evidence-based psychotherapeutic and medication treatments at scale. We may need to integrate specialty behavioral health services more fully into primary care.26 Alternatively, we may need to develop more collaborative and clinically integrated networks with cross-practice support staff, better information, and tracking of referrals. This suggests that screening should be undertaken in the context of an overhaul of the primary care system, including behavioral health integration or all of the supports, such as facilitated referrals and longitudinal registries, that successfully improved delivery of treatment in the adult trials with

FIGURE 3
Histogram of simulation population effects. The distribution of the percent of the maximum population effect achieved by universal screening versus standard assessment in 10,000 simulated populations.
more successful outcomes.\textsuperscript{27} Whether current strategies for doing this would succeed or be sustainable is unclear.

Alternatively, screening should be considered in systems that can efficiently capture the target population for screening and expedite integration with treatment services. For example, perhaps we should screen for adolescent depression in schools.\textsuperscript{28} Again, we recommend that school screening be preceded by an intensive empirical study of the challenges of delivering care when screening occurs in the school setting. For example, there are considerable efficiencies to be gained by school-based screening, such as accessibility to almost all teenagers. However, the teenagers who might not be in school screening (dropouts or truants) may be more vulnerable to depression. Also, it will be critical to verify that schools can provide immediate follow-up and on-site mental health services. Of course, many schools are overwhelmed with their existing tasks and may decline to screen.

**Limitations**

In our simulation, we used a simplified model of care that omits many details of practice and, as such, may represent a best-case scenario. Moreover, the simulation’s relevance depends on having accurate information on how well the primary care system delivers care. Here, the literature is still thin, so the appropriate values for the simulation parameters were uncertain. However, this uncertainty does not undermine our central finding that universal screening would have only a limited effect on population mental health because our simulation’s largest population health effect was \(<20\%\). Even when every system parameter was at the best value observed in practice, the cumulative effect of moderate losses
at each step of the care delivery process resulted in a small population health effect. Another limitation of this study is that we did not consider the direct costs or opportunity costs of screening to practices or patients.

Finally, we did not consider the effects of screening on health equity. The primary care system delivers more and better care to privileged segments of the population than to deprived segments. Therefore, we hypothesize that universal screening works better in well-resourced communities than it does in poor communities. This disparity would be particularly problematic because rates of pediatric mental disorders are higher in deprived populations. In future research on screening or other preventive interventions, researchers should address population variation and disparities in health care system functioning that contribute to worsening inequities in outcomes.

CONCLUSIONS
Screening needs to be evaluated in the context of a realistic model of health care system functioning. Given the current data on health system delivery effectiveness in depression care, there are grounds for concern that a universal depression screening intervention’s population health effects on adolescents will be small. The many points of failure in the mental health care delivery process will limit the population health effects of universal depression screening. We need more research on the details of health care delivery and how it is affected by screening. To achieve better results, we need to identify health care system structures and processes that strengthen the population effectiveness of screening or consider alternative settings in which to conduct it.

ABBREVIATION
FPR: false-positive rate

REFERENCES

adverse health outcomes: a systematic umbrella review. *JAMA Psychiatry.* 2019;76(12):1241–1255


The Potential for Improving the Population Health Effectiveness of Screening: A Simulation Study
William Gardner, Katherine Bevans and Kelly J. Kelleher
Pediatrics 2021;148:s3
DOI: 10.1542/peds.2021-050693C
The Potential for Improving the Population Health Effectiveness of Screening: A Simulation Study
William Gardner, Katherine Bevans and Kelly J. Kelleher
Pediatrics 2021;148;s3
DOI: 10.1542/peds.2021-050693C

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://pediatrics.aappublications.org/content/148/Supplement_1/s3

Data Supplement at:
http://pediatrics.aappublications.org/content/suppl/2021/06/25/peds.2021-050693C.DCSupplemental