


ORIGINAL ARTICLE

Decision curve analysis as a framework to estimate the potential value of screening or other decision-making aids

Michael S. Martin^{1†}  | George A. Wells¹ | Anne G. Crocker^{2‡} | Beth K. Potter¹ | Ian Colman¹

¹School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada

²Department of Psychiatry, McGill University and Douglas Mental Health University Institute Research Centre, Montreal, Quebec, Canada

Correspondence

Michael S Martin, School of Epidemiology and Public Health, University of Ottawa, 340 Laurier Avenue West, Ottawa, ON, K1A 0P9. Email: mmart007@uottawa.ca

Present Address

[†]Mental Health Branch, Correctional Service of Canada

[‡]Department of Psychiatry, Université de Montréal and Institut Philippe-Pinel de Montréal

Funding information

Correctional Service of Canada (CSC)

Abstract

Objectives: There is an increasing debate about the impact of mental health screening. We illustrate the use of a decision making framework that can be applied when there is no sufficient data to support a traditional cost-benefit analysis.

Methods: We conducted secondary analyses of data from 459 male prisoners who were screened upon intake. We compared the potential benefit of different approaches (screening, history taking, and universal interventions) to allocating treatment resources using decision curve analysis.

Results: Screening prisoners for distress at typical levels of sensitivity (75%) and specificity (71%) were estimated to provide the greatest net benefit if between 2 and 5 false positives per detected illness are tolerable. History taking and self-harm screening provide the largest net benefit when only 1 or 2 false positives per detected illness would be tolerable. The benefits of screening were less among those without a recent psychiatric history, ethnic minorities, and those with fewer psychosocial needs.

Conclusions: Although screening has potential to increase detection of treatment, important subgroup differences exist. Greater consideration of responses to positive screens or alternatives to screening are needed to maximize the impact of efforts to improve detection and treatment of mental illness.

KEYWORDS

decision analysis, mental health, prisons, screening

1 | INTRODUCTION

There are high personal, social, and economic costs associated with low rates of detection and treatment of mental illness (Mitchell, Vaze, & Rao, 2009; Senior et al., 2012; Whiteford et al., 2013). A recent simulation model estimated that increasing treatment rates for depression and anxiety disorders by 20–30% would a return up to four dollars for every dollar spent on treatment (Chisholm et al., 2016). Although this analysis did not describe how to achieve this increase in treatment rates, screening is often recommended to improve treatment uptake. However, recent systematic reviews and guidelines found little difference in detection and treatment of mental illness in settings that screen compared to those that do not (Canadian Task Force on Preventive Health Care, 2013; Gilbody, Sheldon, & House, 2008) or no studies of sufficient methodological rigour (Thombs, Ziegelstein, Roseman, Kloda,

& Ioannidis, 2014). Conversely, using indirect evidence that screening tools can identify depression and that treatment is effective for persons with illness, the United States Preventive Services Task Force (USPSTF) recently recommended screening (Siu et al., 2016). Others have questioned this indirect evidence approach arguing that the positive predictive value of screening is too low for clinical use and that newly detected cases benefit less from treatment than those who are detected through routine care (Coyne, Thompson, Palmer, Kagee, & Maunsell, 2000; Mojtabai, 2017; Thombs et al., 2011).

Mojtabai (2017) noted the USPSTF definition of harms focused on side-effects of medications and ignored other harms. For example, false positive screening results cause inconvenience and time/cost of further appointments; those who are incorrectly diagnosed can experience treatment side-effects (Hampton, Daubresse, Chang, Alexander, & Budnitz, 2014; Lilienfeld, 2007) and stigma (Angermeyer, Matschinger,

& Schomerus, 2013). At a system level, inefficient use of resources can result from responding to false positives and overdiagnosis (i.e., treating illness that would have remitted naturally or not led to suffering or impairment; Carter, Degeling, Doust, & Barratt, 2016). Mojtabai described this as an opportunity cost, whereby time and effort devoted to screening may be more profitably used for other activities. Patten's (2002) simulation study found that sustained treatment to prevent relapse could have a greater impact in reducing prevalence of mental illness than increasing access to treatment through activities such as screening.

In high-prevalence settings, the ratio of false positives to true positives is lower (Goldberg, 2014; Mitchell et al., 2009), thus increasing the value of screening. Because inmates have a higher prevalence of mental illness, screening is part of most standards or guidelines (e.g., Livingston, 2009; National Institute for Health and Care Excellence [NICE], 2017) and American courts have indicated screening as a constitutionally guaranteed right for prisoners (American Civil Liberties Union, 2011). The best studied screening tools have sensitivity and specificity values typically in the 65–75% range (Martin, Colman, Simpson, & McKenzie, 2013), considerably higher than routine clinical detection rates of mental illness ranging from 25% to 50% (Jakobowitz et al., 2017; Senior et al., 2012; Teplin, 1990). However, screening is often limited to mental health history taking and/or screening for suicide risk (Dressing & Salize, 2009; Scheyett, Vaughn, & Taylor, 2009), with little or no measurement of current symptoms.

Although benefits of mental health treatment are well studied through meta-analyses, randomized controlled trial (RCTs) often exclude participants who do not meet diagnostic criteria or do not compare outcomes between true positives and false positives (e.g., Yoon, Slade, & Fazel, 2017). Neufeld, Dunn, Jones, Croudace, and Goodyer (2017) recently highlighted the importance of distinguishing true and false positives in treatment effectiveness studies, as they found no association between treatment and depression at follow-up among all service users. However, among the subgroup with diagnoses of depression at the start of the study (i.e., true positives in the screening context), the risk of depression at follow-up was reduced among those who received treatment compared to those who did not.

Although there are no RCTs of mental health screening in prisons, two observational studies have reported 3–5% increases in the proportion of inmates accessing treatment (Evans et al., 2017; Pillai et al., 2016). Neither study looked at duration of care or accuracy of treatment decisions. One prior study examining long-term follow-up to screening found evidence of potential overuse and underuse of treatment resources (Martin et al., 2017). Sixty-nine per cent of inmates who received treatment did not meet diagnostic criteria and 54% of inmates meeting diagnostic criteria received very brief or no treatment. Furthermore, 50% of those with a known history experienced at least one interruption in care of 30 days or longer.

Decision curve analysis (Vickers & Elkin, 2006) can estimate the potential net benefit of screening where the exact harms (i.e., costs of false positives) and benefits (i.e., better outcomes among detected cases) are not fully quantified. In this paper, we apply the method to mental health screening and discuss its potential relevance to policy making and clinical practice.

2 | METHODS

This retrospective cohort study involved secondary analyses of screening data collected as part of routine practice in Canadian prisons and research diagnoses (using the Structured Clinical Interview for DSM Disorders [SCID]) from a prevalence study conducted by the prison service (Beaudette, Power, & Stewart, 2015). Participants were eligible for the prevalence study irrespective of their screening results. Our analyses included all male inmates who completed mental health screening and the SCID. Of 999 inmates who completed screening, 554 were contacted to complete the SCID and 459 (82.9% of those contacted) participated. Participants had a mean age of 35.3. On the basis of self-reported ethnic/racial groups using standardized categories, 60% were White, 25% Aboriginal, 5% Black, Sub-Saharan African or Caribbean, and 10% were other ethnicities sparingly distributed across 24 other ethnic or racial groups. Non-participants were of similar age (mean of 35.9) and ethno-racial distribution (all percentages were within 1%). Inmates who completed the SCID were slightly more likely to be referred for follow-up after screening (33%) than those who did not complete the SCID (30%).

2.1 | Measures and procedure

2.1.1 | Screening

Three self-reported items regarding diagnosis, psychotropic medication prescription(s), or psychiatric hospitalizations just prior to incarceration are used for psychiatric history taking. Inmates also complete the Depression Hopelessness Suicide screening form (DHS; Mills & Kroner, 2004) and the Brief Symptom Inventory (BSI; Derogatis, 1993). The DHS is a 39-item questionnaire designed specifically for offender populations. Five items assess current suicide ideation or plan, recent or multiple prior suicide attempts, or a history of self-harm; a positive response to any of these five items can be used to screen for elevated risk of self-harm (Martin, Dorken, Simpson, McKenzie, & Colman, 2014). A total score and depression and hopelessness sub-scale scores capture the number of endorsed items. An inmate scoring at least 8 on the total score, 7 on the depression sub-scale, or 2 on the hopelessness scale is considered a possible case (Mills & Kroner, 2003). The BSI is a 53-item self-report questionnaire. Nine sub-scale scores and a Global Severity Index are calculated as the average item response. A T-score of 63 or higher (using general adult population norms) on the Global Severity Index or on 2 of the 9 sub-scales is considered a possible case (Derogatis, 1993).

We compared increasingly broad mental health screening strategies that reflect the diverse types of screening of inmates. At each step, we added a new criterion, while retaining prior criteria. The first step was referral of those reporting a recent mental health history. Second, we added self-harm risk based on the five DHS items. Third, we added elevated distress on both the BSI and DHS (i.e., multiple cut-offs). Finally, we added elevated distress on either of the BSI or DHS (i.e., simple cut-offs). The first two steps are similar to jurisdictions that offer less intensive screening (Dressing & Salize, 2009; Scheyett et al., 2009). The third and fourth steps represent symptom-based screening and illustrate the trade-off between reducing

false negatives at the expense of efficiency (the fourth step) and gaining efficiency but missing additional cases (the third step).

2.1.2 | Gold standard diagnostic interview

Inmates were interviewed by a research assistant to complete the SCID for DSM-IV (Biometrics Research Department, 2012) and the modified Global Assessment of Functioning Scale (Hall, 1995). To capture mental illness that caused moderate to severe symptoms or impairment, the case definition was a current diagnosis of a mood, psychotic or anxiety disorder plus a Global Assessment of Functioning Scale score of 60 or less. Diagnoses of substance use from the SCID were used to compare the performance of screening for those with and without a substance use disorder. Research assistants were blind to screening results, and diagnostic interview results were not shared with screening staff.

2.1.3 | Additional file information

We collected information about inmate needs including community functioning prior to arrest, employment, and family history from a semi-structured intake assessment (Brown & Motiuk, 2005). We also collected reintegration potential ratings that are determined on the basis of the results of three structured risk assessments. An inmate is rated as low reintegration potential if they score high on at least two risk assessments, high reintegration potential if they score low on at least two assessments, and moderate otherwise (Correctional Service of Canada, 2014).

2.2 | Analysis

Decision curve analysis (Vickers & Elkin, 2006) estimates the potential diagnostic or prognostic utility of a test based on its operating characteristics (i.e., sensitivity and specificity), the prevalence of the illness, and the relative valuation of the benefits of correctly identifying illness (i.e., true positives) compared to the costs of false identification of a person who is not ill (i.e., false positives). This relative importance of benefits and harms is quantified by the treatment threshold, which can be thought of as an exchange rate (Vickers, Van Calster, &

Steyerberg, 2016). In Table 1, we list examples of harms and benefits of follow-up actions to a positive screen. Although decision curve analysis assumes that false positives represent harms of screening, potential benefits of false positive mental health screening results continue to be debated (for a discussion of benefits and arguments opposing them, see Coyne et al., 2000). We list the potential benefits of false positives as there is no reason that we are aware of that would preclude the consideration of these in weighing the relative importance of true and false positives. We focus on only one harm and one benefit to provide a simplified illustration of the meaning of the treatment threshold. The benefit, experienced by those with true positive results, would be a reduced risk of reoffending due to a reduction in symptoms achieved through appropriate treatment. By contrast, the harm experienced by those with false positive results, would be prolonged incarceration if the offender was treated but showed little sign of improvement (because they had no or little room for improvement). If a decision maker would tolerate no more than four inmates without mental illness having their incarceration inappropriately extended in order to prevent one inmate with mental illness from returning to prison after release, this reflects the belief that costs to society of reoffending are 4 times greater than restrictions on an inmate's liberty due to inappropriately prolonged incarceration (cf. Vickers et al., 2016). In this case, 5 referrals (four false positives) would be tolerable in order to detect one illness, and the treatment threshold would be $1/5 = .20$.

If there are more true positives than weighted false positives, decision curve analysis would recommend treating all those who screen positive to achieve a net benefit. This weighted net benefit for the screened population can be calculated using the formula:

$$Prevalence \times Sensitivity - (1 - specificity) \times (1 - prevalence) \times \frac{Treatment\ Threshold}{1 - Treatment\ Threshold}$$

The first part of the formula (prevalence \times sensitivity) indicates the proportion of the population who would be correctly referred (i.e., true positives). The second half of the formula converts false positives into units that are equivalent to true positives. In nonstatistical terms, true positives are individuals who stand to benefit from treatment. Thus,

TABLE 1 Harms and benefits of triage/assessment or treatment for true and false positives

Ill	Benefit of triage/assessment	Harm of triage/assessment	Benefit of treatment	Harm of treatment
Yes (true positive)	<ul style="list-style-type: none"> Identify needs and plan treatment 	<ul style="list-style-type: none"> None (costs are warranted and part of providing care) 	<ul style="list-style-type: none"> Reduce symptoms Prevent incidents, violence, premature mortality 	<ul style="list-style-type: none"> N/A (risk of adverse effects should be considered and balanced in choosing appropriate treatment)
No (false positive)	<ul style="list-style-type: none"> Potential to propose strategies to address other needs outside of scope of work by mental health staff Document baseline mental health status 	<ul style="list-style-type: none"> Inconvenience Potential stigma Divert resources to assessment that could be used to treat known illness 	<ul style="list-style-type: none"> Prevent later illness among sub-threshold cases Provide support for distressing symptoms 	<ul style="list-style-type: none"> Adverse effects of treatment Abuse of prescriptions in prisoner population Divert treatment resources towards lower need cases who may benefit less Mission creep (e.g., treating non-illness for security reasons) Stigma, including potential prolonged incarceration

we express the net benefit of screening as the proportion of the population who stand to benefit (from intervention that should follow a positive screen). We use the phrase “stand to benefit” because not everyone who is detected by screening will go on to receive treatment (e.g., they may refuse treatment, there may be a waitlist, or they may not benefit from a treatment that is provided).

It is typical in decision curve analysis to compare screening to clinical contact for everyone and for no one (Vickers & Elkin, 2006). The comparison to no clinical intervention illustrates the benefit of screening. The comparison to intervention for everyone compares screening to the option of skipping the screening step and progressing directly to the follow-up service. In the traditional mental health services model, this could be a brief triage for all inmates. Universal public health interventions could also be considered. For example, given that distress is highly common in early incarceration, but may resolve quickly, basic distress tolerance, relaxation techniques, or other coping skills could be taught to inmates in group or self-directed formats that entail low harms (thus a lower treatment threshold might be acceptable).

To interpret a decision curve, a decision maker would identify the range of plausible treatment thresholds (e.g., they might ascertain treatment thresholds from numerous stakeholders and consider the range of values). The strategy that provides the greatest net benefit over this range of thresholds would be recommended. Because our different screening approaches are incremental, we use a single curve to represent the various screening options. We place a point on the curve at the treatment threshold for which it becomes beneficial to use the next least intensive screening. We do so for ease of interpreting the curves (by reducing the number of lines), and for space reasons when presenting sub-group analyses (i.e., to present relevant parameters for each decision curve in tabular instead of a figure for every group).

3 | RESULTS

Traditional accuracy statistics (i.e., sensitivity, specificity and positive [PPV], and negative predictive values) of the screening tests are provided in Table S1. For the overall population, referring an inmate who reports either a recent mental health history, self-harm risk factor, or distress on both the BSI and DHS (i.e., up to the multiple cut-offs criteria) performs comparably to the best studied screening tools among prisoners, with sensitivity of 75% and specificity of 71%.

Figure 1 illustrates the decision curve analysis for the entire population^{2,1} As seen in the figure (and in the first row of Table 2), screening provides the greatest net benefit for treatment thresholds ranging from 0.06 through 0.56. In terms of the intensity of screening that provides the greatest benefit, each strategy is optimal over a relatively narrow range of treatment thresholds. The broadest screening option (i.e., referral for an inmate meeting any of the four criteria) is optimal for thresholds between 0.06 and approximately 0.16 (or approximately 6–16 referrals to correctly detect one illness), with a net benefit for 13 to 17% of the screened population. Requiring an inmate to exceed distress cut-offs on both scales (or to report a recent mental health history or self-harm risk factor) is optimal between a threshold of 0.16 and 0.31 (and approximately 3 to 6 referrals to detect one illness); this provides a net benefit for 7–13% of the screened population. If the

harms of inappropriate referrals are judged to be more significant and no more than 2–3 referrals per correctly identified illness (thresholds of 0.31 to 0.56) are acceptable, restricting screening to history taking and self-harm screening would provide the greatest benefit. However, the maximum net benefit of 7% represents only one third of the prevalence of mental illness. As seen in Table 3, the proportions of individuals with co-occurring substance abuse, self-harm risk, violence while in remand jail for the current sentence, and worse functioning prior to incarceration typically decreased when moving from those with a mental health history to those reporting a self-harm risk factor, and then progressing to those who exceeded both, either and finally no distress cut-off scores. This suggests the benefits of treatment may decrease with increasingly intensive screening efforts.

As seen in Figure 2, when analyses are restricted to those without a recent history, screening provides a smaller net benefit over a lower and narrower range of treatment thresholds (0.06 to 0.39). Screening those without a recent history provides a net benefit for 6–10% of the screened population if the broadest screen is implemented, and if there was a willingness to accept between 6 and 16 referrals per detected case. If 3 to 5 referrals per detected case were tolerable, the largest net benefit—up to a maximum of 5% of those screened—is obtained by referring only those who report elevated distress on both measures, self-harm risk, or a recent history. Thus the net benefit among the group who could be newly detected by screening is approximately half of what was estimated for the entire population (as the net benefit over the same treatment thresholds ranged from 7 to 17% for the full population).

Table 2 provides parameters for the decision curve by various other sub-groups. For space reasons, we have not included the parameters for the treat all case. These can be inferred as optimal thresholds range from 0 to the lowest threshold for the broadest screen, and the benefit ranges from the highest benefit for the broadest screen up to the prevalence of illness. Generally speaking, the benefit of screening was greater for White inmates than those of minority ethnic groups. The net benefit of screening was between 1.5 and 7 times greater for White than non-Aboriginal inmates despite similar prevalence rates in these two groups. When taking prevalence differences into account through relative differences in benefits by dividing the maximum benefit of screening by the maximum benefit of the treat all case (i.e., reflecting primarily the sensitivity of screening), we continued to observe greater benefits for White inmates. Using any screening option other than the broadest screen (i.e., simple cut-offs) would lead to a net benefit of less than half of the prevalence for Aboriginals and less than 1/3 for other minority inmates reflecting lower sensitivity. By comparison, screening with multiple cut-offs achieved 64% of the maximum benefit (.16/.25) and history taking alone would still achieve 1/3 of the maximum benefit for White inmates. There would be no net benefit of history taking for Aboriginal or other ethnic minority inmates. On remaining variables, both the absolute net benefit of screening and the thresholds are generally higher in those groups with a higher prevalence of illness (i.e., there are more inmates with illness who can benefit from treatment both overall and among positive screens); this includes inmates with recent psychiatric histories, substance use disorders, lower reintegration potential and higher family, social functioning and employment needs.

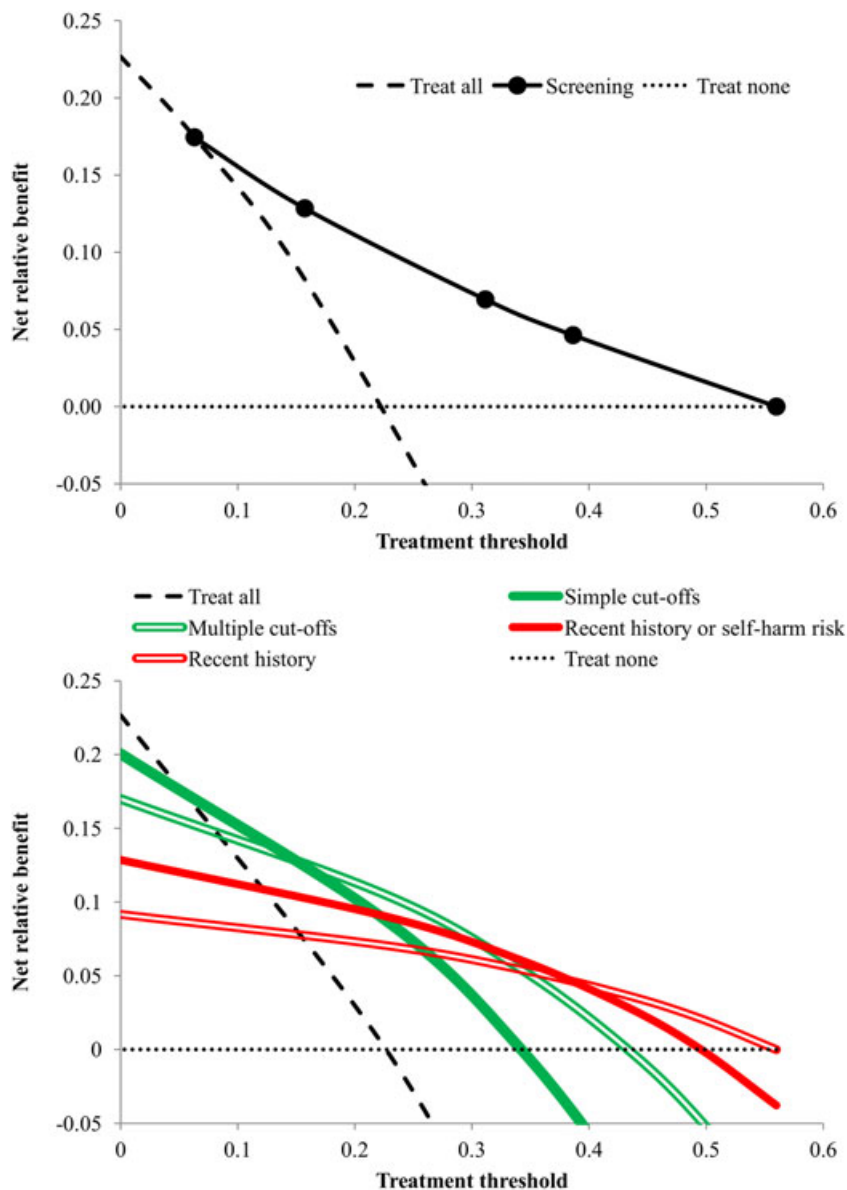


FIGURE 1 Decision curve analysis of screening for the entire inmate population

4 | DISCUSSION

Given the debate about how best to increase access to mental health care, frameworks are needed to support decision making in this area. To our knowledge, this is the first study to apply decision curve analysis to estimate the benefit of screening. Overall, screening appears to be beneficial over a wide range of thresholds. The net benefit of screening with multiple cut-offs which has sensitivity (75%) and specificity (71%) comparable to the best studied tools ranged from 7 to 13% over the range of thresholds for which it was the optimal strategy. However, as discussed in the introduction, observational studies have reported approximately 3 to 5% increases in uptake of treatment following screening (Evans et al., 2017; Pillai et al., 2016), and some of these increases may be through greater provision of services to those who do not require treatment (Martin et al., 2017). Given these gaps from potential to actual impacts of screening, results of our subgroup analyses warrant attention to better understand the limitations of screening as currently practiced.

The finding that the benefit of screening inmates without a recent mental health history is approximately half of the estimate from the full sample emphasizes that spectrum bias (Rice & Thombs, 2016; Thombs et al., 2011) should be addressed by restricting screening studies only to those who could be newly diagnosed, or by conducting stratified analyses. Although there is a benefit of screening individuals with a recent history for treatment thresholds above 0.25 (i.e., tolerating no more than five false positives per true positive), a negative screening result would seem to be insufficient to discontinue mental health interventions that pre-dated incarceration; a lower treatment threshold is likely acceptable for this group of offenders. Given that the recent NICE (2017) guideline for criminal justice populations recommended to screen only those for whom concerns are already noted based on history taking and routine monitoring, this question requires further attention.

Screening also appears to have disparate benefit and treatment thresholds across ethnic groups. Lower sensitivity of mental health history taking among Aboriginal and other minority race inmates is consistent with other studies reporting less access to services in the

TABLE 2 Range of thresholds for which each strategy is most beneficial and corresponding range of net benefits

	n/1000	Prev	Simple cut-offs		Multiple cut-offs		Self-harm		History taking	
			Threshold	Benefit	Threshold	Benefit	Threshold	Benefit	Threshold	Benefit
Total	1000	0.23	0.06, 0.16	0.13, 0.17	0.16, 0.31	0.07, 0.13	0.31, 0.39	0.05, 0.07	0.39, 0.56	0, 0.05
No recent history	837	0.16	0.06, 0.16	0.06, 0.1	0.16, 0.31	0.01, 0.06	0.31, 0.39	0, 0.01	NA	NA
Psychiatric history	163	0.56	0.25, 0.52	0, 0.41	--	--	0.52, 0.67	0.15, 0	NA	NA
Substance use disorder (SUD)	497	0.29	0.09, 0.18	0.16, 0.21	0.18, 0.34	0.09, 0.16	0.34, 0.37	0.08, 0.09	0.37, 0.59	0, 0.08
No SUD	503	0.17	0.04, 0.13	0.09, 0.13	0.13, 0.28	0.05, 0.09	0.28, 0.43	0.01, 0.05	0.43, 0.5	0, 0.01
High RP	269	0.09	0.03, 0.04	0.06, 0.06	0.04, 0.15	0.03, 0.06	0.15, 0.33	0, 0.03	--	--
Moderate RP	456	0.22	0.05, 0.18	0.11, 0.18	0.18, 0.24	0.09, 0.11	0.24, 0.35	0.06, 0.09	0.35, 0.64	0, 0.06
Low RP	275	0.38	0.16, 0.24	0.22, 0.26	0.24, 0.5	0, 0.22	--	--	0.5, 0.59	0, 0
White	610	0.25	0.06, 0.15	0.16, 0.2	0.15, 0.35	0.09, 0.16	0.35, 0.41	0.07, 0.09	0.41, 0.61	0, 0.07
Aboriginal	242	0.23	0.11, 0.14	0.11, 0.13	0.14, 0.26	0.05, 0.11	0.26, 0.35	0.01, 0.05	0.35, 0.4	0, 0.01
Other minority	148	0.13	0.03, 0.21	0.04, 0.11	0.21, 0.25	0.03, 0.04	0.25, 0.5	0, 0.03	--	--
Family functioning need	275	0.37	0.17, 0.23	0.19, 0.24	0.23, 0.42	0, 0.19	--	--	0.42, 0.61	0, 0
No-low family functioning need	725	0.17	0.04, 0.12	0.1, 0.14	0.12, 0.23	0.07, 0.1	0.23, 0.43	0.02, 0.07	0.43, 0.52	0, 0.02
Community functioning need	168	0.34	--	--	0.03, 0.29	0.23, 0	0.29, 0.7	0, 0.23	--	--
No-low community functioning need	832	0.20	0.06, 0.19	0.09, 0.15	0.19, 0.3	0, 0.09	--	--	0.3, 0.52	0, 0
Employment need	540	0.27	0.08, 0.16	0.17, 0.21	0.16, 0.39	0, 0.17	--	--	0.39, 0.63	0, 0
No-low employment need	460	0.17	0.04, 0.15	0.08, 0.13	0.15, 0.23	0.05, 0.08	0.23, 0.4	0.01, 0.05	0.4, 0.44	0, 0.01

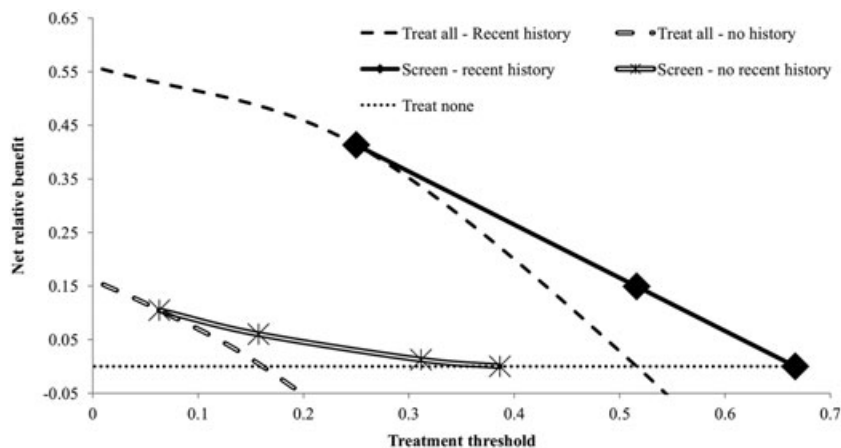
Note. RP = reintegration potential, where higher potential indicates lower risk (or greater likelihood of reintegrating into society).

-- indicates that the screening strategy is never the optimal strategy (i.e., the more sensitive approach, is at least as effective and would thus be preferred so that the maximum benefit is achieved through treating more people who are ill rather than through screening out those who are not).

NA = history taking is not applied within sub-groups of inmates reporting a recent history or not because there is no variation within the groups (i.e., they are defined by this step).

TABLE 3 Needs [95% CI] of individuals in relation to screening risk strata

	Recent history	Self-harm risk	Multiple cut-offs	Simple cut-offs	Screen out
%Moderate-severe impairment	58 [46,70]	57 [42,72]	42 [29,55]	27 [18,36]	17 [12,22]
%Substance use disorder	65 [54,76]	68 [54,82]	48 [35,61]	49 [39,59]	40 [33,47]
%Incidents in remand	19 [10,28]	5 [0,12]	12 [4,20]	7 [2,12]	4 [1,7]
%Employment need	64 [53,75]	66 [52,80]	49 [36,62]	48 [38,58]	52 [45,59]
%Community functioning need	28 [18,38]	20 [8,32]	23 [12,34]	16 [8,24]	10 [6,14]
%Family functioning need	44 [33,55]	36 [22,50]	28 [17,39]	34 [24,44]	16 [11,21]
Health incidents/1000PY	46 [4,89]	9 [0,32]	6 [0,23]	13 [0,33]	2 [0,8]
Violent incidents/1000PY	87 [29,146]	95 [16,174]	92 [26,158]	46 [8,85]	37 [12,62]
Victimization/1000PY	108 [43,173]	95 [16,174]	6 [0,23]	21 [0,47]	20 [1,38]

**FIGURE 2** Decision curve analysis of screening stratified by recent treatment history

community among minority ethnic groups (Prins, Osher, Steadman, Robbins, & Case, 2012; Simpson, Brinded, Fairley, Laidlaw, & Malcolm, 2003). Rather than narrowing pre-existing ethnic disparities in health care, screening could widen them; as such, others have argued unique tools are needed for ethnic and cultural minorities (Ober et al., 2013).

Although consequences of false positives are often considered to be mild, there are often more individuals receiving treatment who do not meet diagnostic criteria than those who do (Martin et al., 2017; Mitchell et al., 2009). Although many of these individuals might have sub-threshold needs or could benefit from preventative services to prevent full blown illness (van Zoonen et al., 2014), in other cases, this may be a wasteful use of resources that can potentially cause harmful side-effects (Hampton et al., 2014; Linden & Schermuly-Haupt, 2014). The negative relationship between broader screening and co-occurring needs and lower maximum benefits and treatment thresholds for those without a psychiatric history are consistent with findings from community studies that individuals detected by screening have lower needs or benefit less from treatment than those who are detected clinically (Cuijpers, van Straten, van Schaik, & Andersson, 2009; Thombs & Ziegelstein, 2013).

These findings highlight some of the challenges of screening and the need to carefully weigh harms and benefits. Following the USPSTF indirect evidence approach, screening is generally recommended on the basis of the understanding that existing treatments are effective but underutilized. In this case, the traditional approach of ascertaining treatment thresholds as described in the methods section could be applied, and a decision maker would evaluate whether screening is the optimal strategy over the range of plausible thresholds (Vickers & Elkin, 2006). However, if screening would be implemented even with equivocal evidence regarding its effectiveness (e.g., in the United States where this is required by law), a model of care would need to be designed around screening in order to maximize its benefits.

Responses following screening could be selected on the basis of their costs and benefits aligning with the treatment thresholds for which screening would be the optimal case-detection strategy. On the basis of the threshold for screening to provide a benefit identified through decision curve analysis, an intervention for which the relative weight of the harms and benefits is below this threshold could be directed by policy. For example, the NICE recommendation of watchful waiting as a first response for mild to moderate mental disorder may be an appropriate first step following broader screening (NICE, 2011). Alternatively, a wider range of lower harm interventions, including self-directed treatment such as bibliotherapy or online (2010) or health promotion and health literacy groups could be considered to minimize potential harms for false positives. In the context of the current findings, inmates who were not receiving psychiatric services prior to incarceration could be offered these lower intensity responses for a fixed timeframe. This timeframe should reflect the time during which symptoms may reflect adjustment issues to prison and may naturally remit (Walker et al., 2014). After adequate follow-up, inmates who continue to display symptoms or who deteriorate, would be offered more intensive interventions such as individual counselling or medication. These more intensive interventions might be considered immediately after screening for those reporting recent histories and/or self-harm risk given that decision makers would likely have a lower

treatment threshold when faced with these pre-existing and higher needs.

Although decision curve analysis can offer valuable insights, it is not without limitation. Because prevalence determines the maximum potential net benefit and positive predictive value determines the treatment threshold at which there would be no effect of the predictive model, decision curves will not generalize to settings with different prevalence of illness. However, as the net benefit formula is straightforward, an interested policy maker could use the sensitivity and specificity of the screening test and prevalence in their setting to create a decision curve for their context.

A second limitation is the tendency of clinicians and patients to overestimate benefit and underestimate harm (Hoffmann & Del Mar, 2015, 2017). Thus randomized clinical trials or other robust study designs are needed to evaluate the actual impact of screening. Following the recommendations of Steyerberg et al. (2013) that a prognostic model should be developed, validated in a replication study and then tested for impact, decision curve analysis can supplement traditional statistics of model performance at the development or validation stage to justify the progression to the subsequent testing step. Many screening tools never progress past the first of these three steps (Martin et al., 2013; NICE, 2011, 2017), increasing the likelihood of overestimating their value. If a decision curve analysis does not support implementation, the costs of a trial would seem unjustified. It is unlikely that decision curve analysis would underestimate the benefit of screening given implementation issues following screening (e.g., patients and/or clinicians deciding treatment is not necessary following a positive screen, and the fact that treatment will not be effective for all those who access it), and because over-estimated benefits and under-estimated harms will result in an underestimate of the true treatment threshold. The results of decision curve analysis could help identify the control condition(s) for an RCT to evaluate the impact of screening. For example, if both screening and the treat all strategy may be likely to result in a net benefit, it would be more appropriate to offer universal intervention rather than treatment as usual (or no screening) as the comparison condition. Alternatively, the results of decision curve analysis could suggest that screening is the optimal strategy, however, it is unclear how intensively to screen. For example, based on the current findings, a decision maker considering a range of thresholds between 0.1 and 0.2 would be faced with the decision between the broadest screen (i.e., referring an inmate reporting any distress for follow-up) and the second most intensive screen (i.e., referring only inmates reporting distress on both measures). An RCT comparing these screening options would be recommended by a decision curve analysis. Because of the risk of inaccurate estimates of treatment thresholds, the third and final step should always be pursued, including a formal cost-benefit analysis that would be given precedence over results of a decision curve analysis.

A final challenge related to ascertaining treatment thresholds is that they may not be generalizable across settings. This is because the intensity of follow-up to a positive screen may vary depending on the priorities of the screening process. For example, screening may aim to identify sub-threshold or prodromal symptoms to prevent onset of illness (Evans et al., 2017; van Zoonen et al., 2014). In this case, the follow-up to screening will be lower intensity and entail less

potential harm; thus, a larger number of false positives would be tolerable and the treatment threshold would be lower. Similarly, if a stepped-care model is in place to limit the intensity of the response to milder symptoms of illness (e.g., using watchful waiting or self-directed therapy as recommended by the National Institute for Health and Clinical Excellence, 2011), this may mitigate the potential harms of overdiagnosis. Conversely, in a resource-limited setting, a higher treatment threshold might be required even if the same follow-up actions are to be provided, due to a greater need to limit over-use by lower need cases.

5 | CONCLUSION

Despite many efforts to improve detection of mental illness through screening, it remains a challenge to balance the costs and benefits of current approaches. Given the limited body of research evaluating the impacts of screening on service use and treatment outcomes, we have shown how decision curve analysis can estimate the potential value of screening. Although there is some early evidence of accuracy and potential net benefits of screening relative to either universal intervention or usual clinical practice, it has yet to be demonstrated that screening is the most cost-effective approach to improve clinical decision-making and mental health outcomes. Guidelines, policies and laws (where they exist) should allow flexibility to consider a wider range of interventions, innovations, and organization of services that ensure that the right care is being provided following screening or in place of it.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

ACKNOWLEDGEMENTS

The first author was supported by a Vanier Canada Graduate Scholarship. Dr Colman is supported by the Canada Research Chairs program. This research was supported by Correctional Service Canada who provided access to data. However, CSC had no role in the conduct of the study. The views expressed are those of the authors and do not necessarily reflect the views of the department or the Government of Canada.

ENDNOTES

¹ To assist in interpreting the tabular presentation of decision curves, we show both a single curve and the traditional decision curve with a separate line for each screening strategy. We have aligned the x-axes to show that the points on the single curve represent thresholds at which a different strategy becomes the optimal one.

ORCID

Michael S. Martin  <http://orcid.org/0000-0002-5056-618X>

REFERENCES

American Civil Liberties Union (2011). Know your rights: Medical, dental and mental health care. Retrieved January 21, 2017, from https://www.aclu.org/sites/default/files/images/asset_upload_file690_25743.pdf

- Angermeyer, M. C., Matschinger, H., & Schomerus, G. (2013). Attitudes towards psychiatric treatment and people with mental illness: Changes over two decades. *The British Journal of Psychiatry*, 203(2), 146–151. <http://doi.org/10.1192/bjpp.112.122978>
- Beaudette, J., Power, J., & Stewart, L. A. (2015). *National prevalence of mental disorders among incoming federally-sentenced men*. Ottawa: Correctional Service of Canada.
- Biometrics Research Department (2012). Structured Clinical Interview for DSM Disorders. Retrieved from www.scid4.org
- Boudreau, R., Moulton, K., & Cunningham, J. (2010). *Self-directed cognitive behavioural therapy for adults with a diagnosis of depression: Systematic review of clinical effectiveness, cost-effectiveness, and guidelines*. Ottawa: Canadian Agency for Drugs and Technologies in Health.
- Brown, S. L., & Motiuk, L. L. (2005). *The Dynamic Factors Identification and Analysis [DFIA] component of the Offender Intake Assessment [OIA] process: A meta-analytic, psychometric and consultative review*. Ottawa: Correctional Service of Canada.
- Canadian Task Force on Preventive Health Care (2013). Recommendations on screening for depression in adults. *CMAJ*, 185(9), 775–782. <http://doi.org/10.1503/cmaj.130403>
- Carter, S. M., Degeling, C., Doust, J., & Barratt, A. (2016). A definition and ethical evaluation of overdiagnosis. *Journal of Medical Ethics, in press*. <http://doi.org/10.1136/medethics-2015-102928>, 42, 705–714.
- Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016). Scaling-up treatment of depression and anxiety: A global return on investment analysis. *The Lancet Psychiatry*, 3(5), 415–424. [http://doi.org/10.1016/S2215-0366\(16\)30024-4](http://doi.org/10.1016/S2215-0366(16)30024-4)
- Correctional Service of Canada (2014). Correctional Planning and Criminal Profile. Retrieved May 1, 2015, from www.csc-scc.gc.ca
- Coyne, J. C., Thompson, R., Palmer, S. C., Kagee, A., & Maunsell, E. (2000). Should we screen for depression? Caveats and potential pitfalls. *Applied and Preventive Psychology*, 9(2), 101–121. [https://doi.org/10.1016/S0962-1849\(00\)80009-8](https://doi.org/10.1016/S0962-1849(00)80009-8)
- Cuijpers, P., van Straten, A., van Schaik, A., & Andersson, G. (2009). Psychological treatment of depression in primary care: A meta-analysis. *British Journal of General Practice*, 59(559), 51–60. <http://doi.org/10.3399/bjgp09X395139>
- Derogatis, L. R. (1993). *Brief Symptom Inventory: Administration, scoring, and procedures manual* (4th ed.). Minneapolis: National Computer Systems.
- Dressing, H., & Salize, H.-J. (2009). Pathways to psychiatric care in European prison systems. *Behavioral Sciences and the Law*, 27(5), 801–810. <http://doi.org/10.1002/bsl.893>
- Evans, C., Forrester, A., Jarrett, M., Huddy, V., Campbell, C. A., Byrne, M., ... Valmaggia, L. (2017). Early detection and early intervention in prison: Improving outcomes and reducing prison returns. *The Journal of Forensic Psychiatry & Psychology*, 28(1), 91–107. <http://doi.org/10.1080/14789949.2016.1261174>
- Gilbody, S., Sheldon, T., & House, A. (2008). Screening and case-finding instruments for depression: A meta-analysis. *CMAJ*, 178(6), 1–11. <http://doi.org/10.1503/cmaj.070281>
- Goldberg, D. (2014). The value of screening in patient populations with high prevalence of a disorder. *BMC Medicine*, 12(14). <http://doi.org/10.1186/1741-7015-12-14>
- Hall, R. C. (1995). Global assessment of functioning. A modified scale. *Psychosomatics*, 36(3), 267–275. [http://doi.org/10.1016/S0033-3182\(95\)71666-8](http://doi.org/10.1016/S0033-3182(95)71666-8)
- Hampton, L. M., Daubresse, M., Chang, H.-Y., Alexander, G. C., & Budnitz, D. S. (2014). Emergency department visits by adults for psychiatric medication adverse events. *JAMA Psychiatry*, 71(9), 1006–1014. <http://doi.org/10.1001/jamapsychiatry.2014.436>
- Hoffmann, T. C., & Del Mar, C. (2015). Patients' expectations of the benefits and harms of treatments, screening, and tests. *JAMA Internal Medicine*, 175(2), 274–286. <http://doi.org/10.1001/jamainternmed.2014.6016>
- Hoffmann, T. C., & Del Mar, C. (2017). Clinicians' expectations of the benefits and harms of treatments, screening, and tests: A systematic

- review. *JAMA Internal Medicine*, 177(3), 407–419. <http://doi.org/10.1001/jamainternmed.2016.8524>
- Jakobowitz, S., Bebbington, P., McKenzie, N., Iveson, R., Duffield, G., Kerr, M., & Killaspy, H. (2017). Assessing needs for psychiatric treatment in prisoners: 2. Met and unmet need. *Social Psychiatry and Psychiatric Epidemiology*, 52(2), 231–240. <http://doi.org/10.1007/s00127-016-1313-5>
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2(1), 53–70. <http://doi.org/10.1111/j.1745-6916.2007.00029.x>
- Linden, M., & Schermuly-Haupt, M. L. (2014). Definition, assessment and rate of psychotherapy side effects. *World Psychiatry*, 13(3), 306–309. <http://doi.org/10.1002/wps.20153>
- Livingston, J. D. (2009). *Mental health and substance use services in correctional settings: A review of minimum standards and best practices*. Vancouver: The International Centre for Criminal Law Reform and Criminal Justice Policy.
- Martin, M. S., Colman, I., Simpson, A. I., & McKenzie, K. (2013). Mental health screening tools in correctional institutions: A systematic review. *BMC Psychiatry*, 13, 275. <http://doi.org/10.1186/1471-244X-13-275>
- Martin, M. S., Dorken, S. K., Simpson, A. I., McKenzie, K., & Colman, I. (2014). The predictive validity of the Depression Hopelessness Suicide screening form for self-injurious or suicidal behaviour among prisoners. *Journal of Forensic Psychiatry & Psychology*, 25(6), 733–747. <http://doi.org/10.1080/14789949.2014.955811>
- Martin, M. S., Potter, B. K., Crocker, A. G., Wells, G. A., Grace, R. M., & Colman, I. (2017). Mental health treatment patterns following screening at intake to prison. *Journal of Consulting and Clinical Psychology*, in press. <http://doi.org/10.1037/ccp0000259>
- Mills, J. F., & Kroner, D. G. (2003). *Depression, hopelessness and suicide screening form: User guide*. Kingston, ON: Author.
- Mills, J. F., & Kroner, D. G. (2004). A new instrument to screen for depression, hopelessness, and suicide in incarcerated offenders. *Psychological Services*, 1(1), 83–91. <http://doi.org/10.1037/1541-1559.1.1.83>
- Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: A meta-analysis. *The Lancet*, 374(9690), 609–619. [http://doi.org/10.1016/S0140-6736\(09\)60879-5](http://doi.org/10.1016/S0140-6736(09)60879-5)
- Mojtabai, R. (2017). Universal depression screening to improve depression outcomes in primary care: Sounds good, but where is the evidence? *Psychiatric Services*, 68(2), appi.ps.2016003. <http://doi.org/10.1176/appi.ps.201600320>, 724–726.
- National Institute for Health and Care Excellence (2017). *Mental health of adults in contact with the criminal justice system*. National Institute for Health and Clinical Excellence.
- National Institute for Health and Clinical Excellence. (2011). Common mental health disorders: The NICE guideline on identification and pathways to care. National Clinical Guideline Number 123. <http://doi.org/clinicalguidelineCG123.2011>
- Neufeld, S. A. S., Dunn, V. J., Jones, P. B., Croudace, T. J., & Goodyer, I. M. (2017). Reduction in adolescent depression after contact with mental health services: A longitudinal cohort study in the UK. *The Lancet Psychiatry*, 4(2), 120–127. [http://doi.org/10.1016/S2215-0366\(17\)30002-0](http://doi.org/10.1016/S2215-0366(17)30002-0)
- Ober, C., Dingle, K., Clavarino, A., Najman, J. M., Alati, R., & Heffernan, E. B. (2013). Validating a screening tool for mental health and substance use risk in an Indigenous prison population. *Drug and Alcohol Review*, 32(6), 611–617. <http://doi.org/10.1111/dar.12063>
- Patten, S. B. (2002). A framework for describing the impact of antidepressant medications on population health status. *Pharmacoepidemiology and Drug Safety*, 11(7), 549–559. <http://doi.org/https://doi.org/10.1002/pds.746>
- Pillai, K., Rouse, P., McKenna, B., Skipworth, J., Cavney, J., Tapsell, R., ... Madell, D. (2016). From positive screen to engagement in treatment: A preliminary study of the impact of a new model of care for prisoners with serious mental illness. *BMC Psychiatry*, 16, 9. <http://doi.org/10.1186/s12888-016-0711-2>
- Prins, S. J., Osher, F. C., Steadman, H. J., Robbins, P. C., & Case, B. (2012). Exploring racial disparities in the brief jail mental health screen. *Criminal Justice and Behavior*, 39(5), 635–645. <http://doi.org/10.1177/0093854811435776>
- Rice, D. B., & Thombs, B. D. (2016). Risk of bias from inclusion of currently diagnosed or treated patients in studies of depression screening tool accuracy: A cross-sectional analysis of recently published primary studies and meta-analyses. *PLoS One*, 11(2), e0150067. <http://doi.org/10.1371/journal.pone.0150067>
- Scheyett, A., Vaughn, J., & Taylor, M. F. (2009). Screening and access to services for individuals with serious mental illnesses in jails. *Community Mental Health Journal*, 45(6), 439–446. <http://doi.org/10.1007/s10597-009-9204-9>
- Senior, J., Birmingham, L., Harty, M. A., Hassan, L., Hayes, A. J., Kendall, K., ... Shaw, J. (2012). Identification and management of prisoners with severe psychiatric illness by specialist mental health services. *Psychological Medicine*, 43(7), 1511–1520. <http://doi.org/10.1017/S0033291712002073>
- Simpson, A. I., Brinded, P. M., Fairley, N., Laidlaw, T. M., & Malcolm, F. (2003). Does ethnicity affect need for mental health service among New Zealand prisoners? *Australia and New Zealand Journal of Psychiatry*, 37(6), 728–734.
- Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., Ebell, M., ... Pignone, M. P. (2016). Screening for depression in adults. *JAMA*, 315(4), 380. <http://doi.org/10.1001/jama.2015.18392-387>.
- Steyerberg, E. W., Moons, K. G. M., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., ... for the PROGRESS Group (2013). Prognosis research strategy (PROGRESS) series 3: Prognostic model research. *PLoS Medicine*, 10(2), e1001381 <http://doi.org/10.1371/journal.pmed.1001381>
- Teplin, L. A. (1990). Detecting disorder: The treatment of mental illness among jail detainees. *Journal of Consulting and Clinical Psychology*, 58(2), 233–236.
- Thombs, B. D., Arthurs, E., El-Baalbaki, G., Meijer, A., Ziegelstein, R. C., & Steele, R. J. (2011). Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: Systematic review. *BMJ*, 343, d4825. <http://doi.org/10.1136/bmj.d4825>
- Thombs, B. D., & Ziegelstein, R. C. (2013). Depression screening in primary care: Why the Canadian task force on preventive health care did the right thing. *Canadian Journal of Psychiatry*, 58(12), 692–696.
- Thombs, B. D., Ziegelstein, R. C., Roseman, M., Kloda, L. A., & Ioannidis, J. P. (2014). There are no randomized controlled trials that support the United States Preventive Services Task Force guideline on screening for depression in primary care: A systematic review. *BMC Medicine*, 12(1), 13. <http://doi.org/10.1186/1741-7015-12-13>
- van Zoonen, K., Buntrock, C., Ebert, D. D., Smit, F., Reynolds, C. F., Beekman, A. T. F., & Cuijpers, P. (2014). Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions. *International Journal of Epidemiology*, 43(2), 318–329. <http://doi.org/10.1093/ije/dyt175>
- Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, 26(6), 565–574. <http://doi.org/10.1177/0272989X06295361>
- Vickers, A. J., Van Calster, B., & Steyerberg, E. W. (2016). Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*, 352, i6. <http://doi.org/10.1136/bmj.i6>
- Walker, J., Illingworth, C., Canning, A., Garner, E., Woolley, J., Taylor, P. J., & Amos, T. (2014). Changes in mental state associated with prison environments: A systematic review. *Acta Psychiatrica Scandinavica*, 129(6), 427–436. <http://doi.org/10.1111/acps.12221>
- Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., ... Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, 382(9904), 1575–1586. [http://doi.org/10.1016/S0140-6736\(13\)61611-6](http://doi.org/10.1016/S0140-6736(13)61611-6)

Yoon, I. A., Slade, K., & Fazel, S. (2017). Outcomes of psychological therapies for prisoners with mental health problems: A systematic review and meta- analysis. *Journal of Consulting and Clinical Psychology*, *in press*. <http://doi.org/10.1037/ccp0000214>, 85, 783–802.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Martin MS, Wells GA, Crocker AG, Potter BK, Colman I. Decision curve analysis as a framework to estimate the potential value of screening or other decision-making aids. *Int J Methods Psychiatr Res*. 2018;27:e1601. <https://doi.org/10.1002/mpr.1601>