Significantly off target: A commentary on Fazel et al.'s (2024) umbrella review of research on the Risk-Need-Responsivity framework.

James McGuire

University of Liverpool, UK

Michael C. Seto

University of Ottawa, Canada

Jason Davies

University of Swansea, UK

Friedrich Lösel

University of Erlangen-Nürnberg, Germany, and

Institute of Criminology, Cambridge, UK

Significantly off target: A commentary on Fazel et al.'s (2024) umbrella review of research on the Risk-Need-Responsivity framework.

Abstract

We write in relation to a review article concerning the status of the Risk-Need-Responsivity (RNR) model of offender rehabilitation that is widely utilised in criminal justice services. The article to which we refer, by Fazel et al. (2024), was published in the *Journal of Criminal Justice*. We make a number of observations on the work reported in the paper, which together indicate that the conclusions drawn by its authors are not scientifically justified and are misleading.

We offer this critical position based on a detailed scrutiny of the report by Fazel et al. and the manner in which they conducted their review. We find that there are several features of their work that make the conclusions they arrived at highly questionable. They comprise (a) a dismissive approach to what are in fact mainly positive outcomes; (b) an inappropriately narrow selection procedure which led to the exclusion of many relevant meta-analyses; and (c) the use of unsuitable and misleading methods of systematic review quality appraisal. Overall, we conclude that there is a misrepresentation of the RNR model at a theoretical level which has led to a misjudgement of both its heuristic and its practical value. We agree that there are weaknesses in portions of the relevant research on RNR, and we agree that the model is not an entirely comprehensive account of the criminal justice rehabilitation enterprise. While we appreciate some of the critical comments made by Fazel et al., their main conclusions are unwarranted and we reject their suggestion that RNR should not provide a basis for design and delivery of rehabilitation services in new jurisdictions.

Significantly off target: A commentary on Fazel et al.'s (2024) umbrella review of research on the Risk-Need-Responsivity framework.

Writing recently in the *Journal of Criminal Justice*, Fazel, Hurton, Burghart, DeLisi and Yu (2024) have reported an umbrella review of what many experts have regarded as the best-established and most widely applied approach to, and framework for, the process of rehabilitation in criminal justice: the Risk-Need-Responsivity (RNR) model described by Bonta and Andrews (2024; the seventh edition of a work first published in 1994). Prior to addressing these issues in detail, we wish to indicate first that we have not been involved in the development of RNR and have independent views on the topic developed over many years.[1]

The most recent version of the RNR model entails a series of 18 principles, but in its most widespread application particular attention is focused on the three from which its name is derived. The *risk* principle means that intensity of service should be matched to the risk level of cases, and interventions should concentrate mainly on moderate and higher levels of risk. The *need* principle entails the assessment and targeting of predominant criminogenic needs, that is, changeable factors that are associated with criminal activity, rather than other needs that the individual may have. The *responsivity* principle has two components: *general* responsivity that advocates the use of behavioural, social learning and skill-building strategies of change; and *specific* responsivity which requires adapting modes of service delivery to settings and to the assessed characteristics of individual participants (such as age, gender, ethnicity, literacy level, strengths).

In their review, Fazel et al. (2024) conducted a literature search for meta-analyses of the empirical grounding that informs the principles underpinning the RNR model. Their report covers a 20-year period and includes studies published between 01/01/2002 and 15/12/2022.

---

[1] RNR developers and others have published their own responses to Fazel et al. (2024) in the August 2024 issue of the newsletter of the International Association for Correctional and Forensic Psychology (IACFP). See: https://www.myiacfp.org/the-iacfp-newsletter/

Following specified procedures, Fazel et al. searched the PubMed, PsycNet, and Scopus databases, and the Cochrane Library of systematic reviews, plus three sources of information on grey literature. They identified a total of 26 meta-analyses and systematic reviews that were considered to provide evaluative information on the RNR model. Some reviews addressed more than one of the model's principles, and collectively they subsumed more than 450 primary studies. The umbrella review identified 7 reviews of the risk principle, 6 of the need principle, 15 of the general responsivity principle, and 9 of the specific responsivity principle. They applied measures of study quality to the retained reviews, and summarised the findings from them in a series of statistical tables. Their review has also been published in briefer form in a professional journal (Fazel & Favril, 2024).

Fazel et al. (2024) considered that the evidence supporting the RNR model was mixed and often of poor quality. Their principal conclusion (p. 8) was that "despite RNR's widespread use in criminal justice and claims from experts, we found that the evidence base is mostly low quality and inconsistent". They considered (p. 5) that "the findings on effect sizes and low quality of the underlying evidence raise important and timely questions regarding the continued application and utility of RNR as a model informing criminal justice services", and as things currently stand (p. 8) "introducing RNR into new jurisdictions should not be recommended".

These are very strong positions that require a very sound empirical basis. However, in our view, it would be a serious mistake if those conclusions were to become uncritically accepted. Fuller examination, reported below, highlights several major issues with the Fazel et al. (2024) review. The cumulative effect of them is a very marked misrepresentation of the scientific status and practical utility of the RNR model and of the research associated with it.

Theoretical background

To establish the context, it is critical to recognise the RNR model's theoretical basis, that of the General Personality and Cognitive Social Learning (GPCSL) theory and its yet broader framework of the Psychology of Criminal Conduct (PCC). Whilst this framework and wider

theory is acknowledged by Fazel et al. at one point in their article, they overlook the extensive links between GPCSL theory and other criminological theories, such as control, subcultural, strain, differential association, and related concepts, to which extensive attention is given in the successive editions of the work of Bonta and Andrews, and in other publications which they, their colleagues, or other researchers in this field have produced.

The authors have to some extent taken RNR out of its GPSCL and PCC context: as just noted they make only one reference to the former, an approach that now entails 18 principles (in the 2017 edition of the Bonta and Andrews book there were 15). While we agree that the risk, need, and responsivity principles remain the most prominent and frequently tested of them, other elements have been added as the model has developed over time and as evidence has accumulated. Consequently there are overarching principles that are regarded as being as important as risk, need, and responsivity. However the latter three have been preserved in successive iterations and are considered to be the core of GPSCL.

In addition to this lack of theoretical context, we have noted three major sets of problems within the paper by Fazel et al. (2024) and in what follows we address each in turn.

Issue 1: Negative conclusions overstated and mostly unfounded

The first and most obvious difficulty presented by the Fazel et al. (2024) review is their critical conclusions concerning the status of the RNR model, despite the majority of the aggregated effect sizes quoted in their paper providing positive support for it. In a series of result summaries (Figures 2-6), they show a total of 32 odds ratios. Among those, all the mean odds ratios (ORs) are in favour of RNR principles, and 24 of them have confidence limits that exclude the null effect (an OR of 1.0). Three reviews have a lower limit at OR = 1 and four reviews had confidence limits extending as low as 0.8 (two reviews) or 0.7 (also two).

Risk principle. With regard to the risk principle, seven reviews were identified. The majority of them (5/7) had confidence intervals that did not include the null effect. The authors state (p. 5) that "around half the effect sizes were not significant for the risk principle and the

impact of criminogenic needs". They reached that conclusion after eliminating three of the aforementioned seven reviews because of suspected author bias, based on the rationale that the authors were RNR developers or allies, without examining whether there might be a bias by comparing the excluded and retained reviews. Their tabulation of the remaining four reviews does not provide a sound basis for the conclusion they draw. An important finding reported by Landenberger and Lipsey (2005) concerned the relationship between treatment effect size and participants' rated risk of recidivism. They obtained a significant beta coefficient of 0.27 between those two variables (shown in their Table 5, p. 25).[2] That is strong support for the risk principle.

Need principle. The approach which Fazel et al. (2024) adopt to evaluation of the need principle is particularly problematic, with very few reviews identified that are closely relevant to what that principle entails. Their operationalisation of it defines it almost entirely in terms of risk, hence it becomes a second test of the risk principle rather than criminogenic needs.

First, they cite results from one of their own studies (Fazel et al., 2022), focused on three risk-related measures, the *Level of Service Inventory, Revised* (LSI-R, developed by the main RNR authors), the *Static-99* (Revised), and the *Psychopathy Check List - Revised* (PCL:R), using the Area Under the Curve (AUC) as the outcome indicator, as a measure of predictive power in relation to recidivism. This is a strange test of the need principle, which suggests that interventions that target criminogenic needs are more effective than those that do not. As the name implies, the Static-99(R) is a static risk assessment measure, based on factors that cannot change (e.g., criminal history, prior sexual victim characteristics). No intervention can change criminal history or prior sexual victim characteristics. In contrast, interventions can address self-regulation deficits, attitudes, interpersonal skills, and other changeable factors assessed using the *STABLE-2007*, a dynamic risk measure that is often used in conjunction with the Static-

---

[2] In Fazel et al.'s Figure 2, it is not entirely clear how the odds ratio of 2.8 is related to this, or why it is presented in the way that it is, as it is not reported by Landenberger and Lipsey (2005) or Lipsey et al. (2007) with any confidence limits. Fazel et al. may however have obtained that information from the meta-analysis authors.

99(R). Yet Fazel et al. (2024) did not include the meta-analysis reported by Brankley, Babchishin, and Hanson (2021), even though this systematic review falls within their search criteria.

The same problem applies partially to the LSI-R and PCL-R. Many of the items on these risk measures are also historical or static in nature, including items about criminal history, previous problems in work or education, and prior substance misuse. Some of the LSI-R items are dynamic and could therefore be relevant to an analysis testing the need principle if they are considered separately, for example, looking at changes in pro-criminal attitudes and beliefs. The PCL-R includes dynamic information, but the scoring is based on appraisal of the person's lifetime, so any changes would have to be sustained for a substantial period of time to affect scores.

Second, the evidence regarding the need principle relies on single AUCs as an index of predictive accuracy. This index does not contain specific information about criminogenic needs, and is not a direct test of the need principle either since it is *change* on measures of criminogenic needs that are essential. Evidence concerning *criminogenic needs* as defined within the RNR model is essential in allocating individuals to appropriate interventions and services. That cannot be done on the basis of an overall summary score. It should instead be grounded in a profile of criminogenic areas, such as substance abuse, antisocial attitudes, or a network of criminal associations; with assessment covering those and then being used to inform the kinds of services delivered.

Third, Fazel et al. (2024) unquestioningly include their own reviews in this analysis, having made an argument to exclude studies previously on the basis of potential author allegiance effects. It is noted that there is no examination of whether they might show an allegiance effect to their own reviews (e.g., in rating review quality) compared to the independent reviews they included, yet they pre-emptively excluded reviews involving RNR developers.

Responsivity principle. For general responsivity, 4 of the 14 reviews (one of the 15 reviews that they found was pre-emptively excluded in case of allegiance bias) reported a null

effect but it is unclear why this these outweigh findings from the remaining 10 to carry such negative implications regarding the status of the RNR model. For specific responsivity, none of the 9 meta-analyses that were included showed a null effect, indicating robust evidence for the specific responsivity principle.

There are some incongruities observed when individual reviews are examined. We illustrate this with reference to the review of cognitive-behavioural therapy (CBT) interventions by Landenberger and Lipsey (2005; Lipsey, Landenberger & Wilson, 2007). We focus on this because CBT or its variants have been considered to come closest to interventions that adhere to the *general responsivity* principle within RNR. Neither of the aforementioned reviews is included in Fazel et al.'s (2024) reference lists, but we infer that they are the correct sources.

As accurately reported in Figure 5 of Fazel et al. (2024), the overall outcome that Lipsey et al. (2007) cite for the 58 studies included in their review, as an indicator of observed treatment effect, is a mean odds ratio of 1.53 at an average time-point 12 months after intervention. That corresponds to a 25% decrease in reoffending in treatment groups compared to controls. When interventions with "best practice" CBT features were isolated, "compared to a control group recidivism of .40 (the overall mean), this represents a decrease to a recidivism rate of .19 in the treatment group, that is, a 52% decrease" (p.20). Thus "the most effective configurations of CBT produced odds ratios nearly twice as large as the mean" (p.21). In addition, there were no significant differences found between randomized experiments, of which 19 were included in Lipsey et al.'s review (33% of their study sample), and nonrandomized designs. The nature and extent of these patterns and variations is not reflected in the methodology applied in Fazel et al.'s umbrella review. Again, we should not dismiss a set of studies as not supporting a hypothesis because the confidence limits of the means odds ratio include the null effect, if the studies which generated lower effect sizes have fewer of the features predicted to have a positive association with a successful outcome. If such a pattern of results emerges, it favours the treatment hypothesis rather than detracting from it.

8

The conclusion reached by Lipsey et al. (2007) is similar to the one later drawn by Wilson (2016, p.210): "Overall, the empirical evidence for the effectiveness of CBT programs is relatively strong with numerous high-quality randomized trials demonstrating positive results across a diversity of approaches and therapeutic elements. This is consistent with the general recommendation of the RNR model to use cognitive-behaviorally based methods". That position is markedly at odds with the one forwarded by Fazel et al. (2024) in their paper. The findings of Lipsey et al. (2007) and Wilson (2016) may of course draw a rather positive picture of the effectiveness of correctional treatment. Effects are often smaller, particularly in custody (e.g., Koehler & Lösel, 2025).

A recommendation for utilising CBT or its variants does not imply that other approaches may not also be effective or that CBT is presumed by the current authors to be generally superior to any others: simply that other approaches have been subject to less testing within criminal justice settings. However the CBT approach is commensurate with the assumptions of RNR in theoretical terms (the GPCSL). CBT and its variants constitute the most widely examined approach to individual change in criminal justice settings. Other approaches, for example employment training, therapeutic communities and other psychotherapeutic approaches, could also be utilised; but so far they have not benefitted from the same level of empirical support (Weisburd, Farrington & Gill, 2016). Future evaluative research may find that other approaches are equally or potentially more effective. CBT is unlikely to be the only approach to yield positive outcomes, given that several other types of therapy also produce fundamental alterations such as the development of new neural connections (McGuire, 2025).

Multiple principles. Even when some effect sizes on one of the three RNR elements are small or not statistically significant, they can be relevant when not seen in isolation. Various meta-analyses have shown that programmes have much stronger effects on the outcome of recidivism, when all three RNR principles are operating (e.g. Bonta & Andrews, 2017, 2024); hence, the cumulative or synergistic effect of the principles operating together is a more appropriate test of the model. For example, evidence in support of the need principle still

depends on interventions being appropriately titrated to risk level (there is no room to show differences if all participants are very low risk to reoffend) and interventions being responsive to users (a highly competent program with written exercises and workbooks still falls short if delivered without adaptation to individuals who cannot read well). This point is illustrated in Figure 1, adapted from Bonta and Andrews (2017). The figure shows the respective effect sizes obtained when we compare the effects of different levels of adherence to what is specified in the model. The same pattern of differential effects for different numbers of fulfilled RNR criteria has been found in meta-analyses on the treatment of sexual offenders (Hanson et al., 2009), juvenile offenders (Koehler et al., 2013), and the comparison of human service versus deterrence-based programs (Lowenkamp et al., 2010).

< Insert Figure 1 about here >

Issue 2: Study selection

A major deficiency of the Fazel et al. review, which significantly compromises the authors' approach and leads to the mistaken conclusions they draw, is the very limited strategy they employed in the selection of literature for inclusion. Not all studies or reviews with a direct and pertinent bearing on the status of the RNR model necessarily contain the words *risk*, *need*, or *responsivity* with in the search fields used. Numerous studies that are reviews of aspects of the model are likely to be missed if attention focuses exclusively on those words as search terms.

For example, a review by Lowenkamp et al. (2010), which employed meta-analytic techniques without labelling itself as a meta-analysis (p. 371), found support for the RNR "program philosophy". This review tested aspects of the RNR model as applied through a wide range of programs in delivery of criminal justice services in the state of Ohio.[3] Other important missing reviews include the work of Dowden, Antonowicz and Andrews (2003) on the effectiveness of relapse prevention; of Wilson, Bouffard and MacKenzie (2005) on group-based

---

[3] Two related studies, though not meta-analyses, also found clear support across a range of program sites (Lowenkamp, Latessa & Hollinger, 2006; Lowenkamp, Latessa & Smith, 2006).

CBT; of Aos, Miller and Drake (2006) on the comparative cost-effectiveness of different criminal justice interventions; of Hollis's (2007) evaluative report on the delivery of several offending programs in England and Wales; of Lösel and Schmucker's (2005; Schmucker & Lösel, 2015) two meta-analyses of the effectiveness of sexual offending interventions, the latter including only primary studies with scores of 3-5 on the Maryland Scale of study quality; of a review by Stewart, Gabora, Kropp and Lee (2014) on family violence; and of a meta-analysis by Gannon, Olver, Mallion and James (2019) who reported a large-scale review of specialized psychological treatments. We add that even by 2012, there were already over 100 meta-analytic reviews of criminal justice interventions (many more have been published since then), and the RNR principles emerged with substantial empirical support (McGuire, 2013).

Several of those reviews are grounded primarily in evaluations of the effects of programme delivery in conditions of routine practice. Each of them found support for elements of the RNR model, even where that was not an explicit part of their objectives. Most incorporated methods of appraising the quality of the studies they retained. Though they have a close relevance to the status of the RNR model, due to the narrowness of search terms employed they were not located or retained for analysis in the paper by Fazel et al. (2024).

Issue 3: Quality appraisal and testing of bias

We agree with Fazel et al. (2024) that portions of the evidence for the RNR model are weak. The proportion of studies employing RCT methodology remains disappointingly low. In some non-randomized studies and the reviews in which they are subsumed, there is uncertainty over whether experimental and control groups are equivalent in crucial ways; and too often there are substantial proportions of non-starters and dropouts.

Also, while calling for better-controlled evaluations, it is important not to be naïve about what is feasible in penal and justice services. For legal and practical reasons, it is often not possible to allocate individuals with serious offenses to different levels or types of service or a no-treatment control group for hypothesis-testing purposes (an issue recently addressed by

Koehler & Lösel, 2025). Even advocates of true experiments in criminology such as the Academy of Experimental Criminology state that, while RCTs should be applied as much as possible where it is feasible and justified, they should not be seen as the one and only gold standard (Farrington et al., 2020).

It is sometimes preferable to conduct *practical trials* that provide opportunities for testing of hypotheses in routine service settings. Many criminal justice evaluations are "real world" practical trials, reflecting everyday service delivery. Despite their limitations, the results of practical trials may be potentially more generalisable, as they contain the "noise" typically found in everyday service settings.

The application of the *AMSTAR-2*. The issue of poor study quality is not as widespread as Fazel et al. appear to portray it. Half of the reviews retained for the analysis of the risk principle achieved a satisfactory *AMSTAR-2* rating (classified by Fazel et al. as a score of 8 or above). The *AMSTAR-2* is a 16-item checklist and rating scale of the scientific quality of a systematic review (Shea et al., 2017). However only a small proportion of its constituent items directly focus on meta-analytic methodology as such, and most reviews on RNR issues are meta-analytic. Among the 14 general responsivity studies, ten have a score above the required score of 8; the remaining four are only marginally below it (scores of 7 or 7.5). All but one reaches the required *ROBIS* rating of 2/4 and six have scores of 4/4. Obtained scores are not good for specific responsivity, but the nine sets of scores shown are based on evaluation of just four reviews: among them, four of the scores are based on one of the reviews and three on another. Clearly this pattern is fairly mixed but we would question whether it leads to a conclusion as negative as that drawn by Fazel et al.

Thus the authors' inferences are somewhat at odds with the requirements they stipulate for design quality. The majority of the included reviews meet or surpass the score threshold they set for acceptability on the *AMSTAR-2*. However, this leads to a further issue concerning the way the *AMSTAR-2* is applied; the developers of the instrument state that it is "not intended" to be

used as a single score of quality assurance (Shea et al., 2017, p.1). Despite this, a total score is the only information provided by Fazel et al. (2024) for the included reviews.

Additional quality criteria: In addition to global ratings based on existing tools, Fazel et al. (2024) considered the following criteria as evidence of quality: Total sample size exceeded 1,000; No evidence of excess statistical significance; Prediction interval excluded OR = 1.0; I^2 (a measure of heterogeneity in effects) was less than 50%; and an absence of publication bias. The rationale for these is not provided, and these specifications should not be accepted without critical consideration. For example, the sample size requirement appears both arbitrary, and may be unrealistic (even for a meta-analysis) in the practice of offender treatment; and it is a very limited proxy of quality. This has implications for the analyses of excess statistical significance which should be employed very cautiously as it may be misleading to rely on a single test for this without examining the actual coverage of a review in detail (Ioannidis, 2013). Further, the largest study included in a review can only be considered the most accurate estimate of effect if it is a well-conducted piece of work; sample size in itself is not a sufficient indicator of study quality and cannot be taken to imply that the study can be used as an anchor compared to others.

Authorship bias: The typical case in much research is one in which the researchers who originate a theory or model also conduct primary studies on it. It is probably unrealistic to expect otherwise, given how expertise is acquired and developed, and how research is funded and implemented. Typically, researchers do not constantly migrate between unconnected areas of work in directions that entirely avoid focusing on the same scientific question. If they did, there might never be any establishment of laboratories, centres of excellence, or research departments focused on specific problems.

Hence, researchers working on the same topics over extended periods is the normal state of affairs. In the Fazel et al. paper, authorship bias is assumed to be present; many reports and papers with RNR authors' or their affiliates' names on them are thus excluded from review. However, affiliation does not automatically mean author bias in favour of results that support

their own positions. Undoubtedly that can occur, and it may have an influence on the outcome (Petrosino & Soydan, 2005): but it has to be investigated and demonstrated, rather than presumed or implied. Furthermore, if those who have propounded a theory find more support for it than other researchers, that is not necessarily because of bias: it could be because they performed the work more carefully than colleagues working elsewhere. This may not be a strong argument, but we suggest that there should be empirical grounds for excluding papers from a review rather than a presumption that bias is inevitable, as Fazel et al. (2024) have done in relation to risk (as discussed earlier).

None of this is intended to suggest that bias does not exist, but that it needs to be probed rather than assumed. "Independent" evaluators are also no guarantee for full neutrality, for example when they have developed their own competing theories or programs. Maybe hostile authors should also be excluded as they may be biased in the opposite direction to those whose allegiance is (apparently) established. The approach to quality appraisal which Fazel et al. (2024) employ has been used in other umbrella reviews: but the only sources they cite are papers written by themselves. Should this then be regarded as an instance of author bias?

Closing issues

One of the dangers of umbrella reviews is the risk of over-simplification. We acknowledge and commend that the first author of the paper under discussion here is a proven expert on umbrella reviews in various fields, and the research team has considerable experience of this kind of work. Therefore, we wonder why the article does not contain a paragraph on limitations of this method, for example on the overlap of primary studies in various meta-analyses on the same topic, which reduces data independence (Koehler & Lösel, 2025). While Fazel et al. indicate those reviews which subsume overlapping studies, they do not offer information on the extent of this or how far it may affect the conclusions drawn.

Concerning this, there is a level above which evidence cannot be properly synthesized. The end result may be that important aspects of studies are not analysed, with the outcome then

being a superficial impression of what has been found. There is as yet no all-encompassing, ubiquitously endorsed theory within criminology, but a series of models that apply to different levels of explanation dependent on the scale or context of what is under consideration. Similarly, there is no single or simple answer to a complex question regarding the relationship between interventions and outcomes inside service delivery systems as multifaceted as the criminal justice systems, with populations as variegated as those contained within this system. Umbrella reviews can provide some orientation on such complex topics, but we should also recognize their limitations.

Some of the reviews summarized in Figure 2 of Fazel et al'.s paper should not automatically be regarded as *not* supporting the RNR model because they contain studies with low or zero effects. Analysis of moderators within a review can distinguish reasons for heterogeneity, and a null or negative effect may be a product of any number of such variables. This is a problem that makes the findings of an umbrella review difficult to interpret, and such a collection of outcomes needs to be investigated more thoroughly (for example by examining moderators).

On the basis of the observed pattern of results and the conclusions drawn from them, we echo a comment made by Polaschek (2024). Imagine that these authors were to present the series of findings contained in their paper at a cancer conference, "and then to suggest that we stop implementing the treatment of said cancer until we 'know more'. That would immediately raise the important 'compared to what?' question (i.e., what are the effects of returning to the status quo while we wait?)." Alongside this we note that the mean effect sizes of several widely implemented medical treatments are themselves very variable and sometimes include the null effect: for example of chemotherapy for cancer (Wise, 2016). Yet who would refuse to deliver such treatment even if its effects were very small?

We conclude with two final remarks. First, on page 7 of their paper, Fazel et al. (2024) reprimand the authors of other reviews for not including references to works they have cited. Unfortunately they do exactly the same themselves, indeed quite extensively. Of the studies cited

either in the main text of their article, or in the supplementary materials, we counted no fewer than 15 references either in the article or in the supplementary materials for which no source details are given. As with some of the other criticisms they make, the faults the authors allegedly find elsewhere are also present in their own paper. Such factors are important to note but may simply signify human error rather than deliberate obfuscation or an attempt to mislead.

Second, Fazel et al. (2024) attribute the "popularity" of RNR -- by which we believe they mean the extent to which it has been adopted within criminal justice services in many places -- to statements made about its underlying evidence base, because it is sometimes "contrasted to newer models with less well-developed research in support" (p.2) such as the Good Lives Model (GLM). This not an especially apt comparison.[4] The GLM can scarcely be described as new: it was proposed by Ward and Stewart in 2003, over 20 years ago. Even so there is very little evidence of positive outcomes for programs using the Good Lives Model (Netto, Carter & Bonell, 2014; Willis & Ward, 2013, 2025; Zeccolla et al., 2021); and its basic assumptions raise serious difficulties about its scientific status (Serie et al., 2021; Mallion, Wood & Mallion, 2020). The GLM has so many variables it runs into the error known in the philosophy of science as the under-determination of theory by data (McGuire, 2025). Placing it alongside RNR as a comparator or competitor is far from convincing, as the value of it as a stand-alone or complementary intervention remains unproven.

Despite the criticisms of Fazel et al. (2024), and given the reservations expressed in this paper, RNR remains the most soundly-based and useful guide available in the field of criminal justice rehabilitation. Our net conclusion having considered their umbrella review is that it contains omissions, errors, and misinterpretations, and that the markedly negative conclusion it offers concerning the RNR model is unfounded. We fear that its impact on correctional rehabilitation outcomes will be deleterious and that it will cause unnecessary confusion in this area of practice and policy.

---

[4] The present paper is not a critique of the GLM and it is not the focus of our comments, but we felt it required at least a few sentences because Fazel et al. mentioned it as perhaps more promising than RNR.
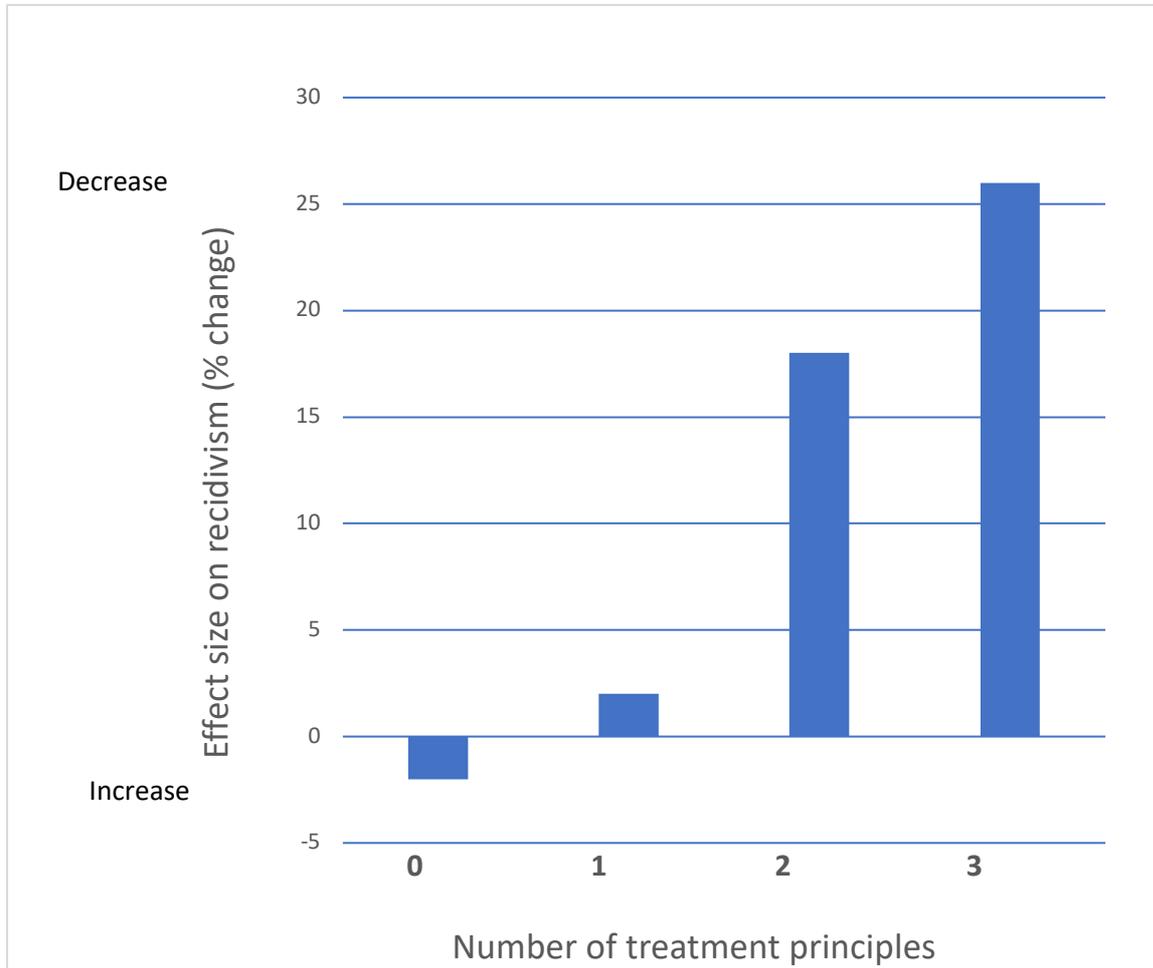
Figure 1: Mean effect size by adherence to RNR principles. Adapted from Bonta and Andrews (2017).

References

Aos, S., Miller, M. & Drake, E. (2006). *Evidence-Based Adult Corrections Programs: What Works and What Does Not*. Olympia, WA: Washington State Institute for Public Policy.

Bonta, J. & Andrews, D. A. (2017). *The Psychology of Criminal Conduct*. 6th edition. New York, NY & Abingdon: Routledge.

Bonta, J. & Andrews, D. A. (2024). *The Psychology of Criminal Conduct*. 7th edition. New York, NY & Abingdon: Routledge.

Brankley, A. E., Babchishin, K. M. & Hanson, R. K. (2021). STABLE-2007 demonstrates predictive and incremental validity in assessing risk-relevant propensities for sexual offending: A meta-analysis. *Sexual Abuse*, *33*(1), 34-62.

Dowden, C., Antonowicz, D. & Andrews, D. A. (2003). The effectiveness of relapse prevention with offenders: a meta-analysis. *International Journal of Offender Therapy and Comparative Criminology*, *47*(5), 516-528. DOI: 10.1177/0306624X03253018

Farrington, D. P., Lösel, F., Braga, A. A., Mazerolle, L., Raine, A., Sherman, L. W. & Weisburd, D. (2020). Experimental criminology: Looking back and forward on the 20th Anniversary of the Academy of Experimental Criminology. *Journal of Experimental Criminology, 16*, 649-673. DOI: 10.1007/s11292-019-09384.

Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. *Journal of Criminal Justice, 81*, Article 101902. DOI: 10.1016/j.jcrimjus.2022.101902

Fazel, S. & Favril, L. (2024). Results from an umbrella review of the evidence underpinning RNR principles: inconsistent and poor quality. *Probation Quarterly*, *33*, 9-12.

Fazel, S., Hurton, C., Burghart, M. DeLisi, M. & Yu, R. (2024). An updated evidence synthesis on the Risk-Need-Responsivity (RNR) model: Umbrella review and commentary. *Journal of Criminal Justice*, 92, DOI: 10.1016/j.jcrimjus.2024.102197

Gannon, T. A., Olver, M. E., Mallion, J. S. & James, M. (2019). Does specialized psychological treatment for offending reduce recidivism? A meta-analysis examining staff and program variables as predictors of treatment effectiveness. *Clinical Psychology Review*, *73*, DOI: 10.1016/j.cpr.2019.101752

Hanson, R. K., Bourgon, G., Helmus, L. & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior, 36*(9), 865–891. DOI: 10.1177/ 0093854809338545

Hollis, V. (2007). *Reconviction Analysis of Interim Accredited Programmes Software (IAPS) data*. London: Research Development Statistics, National Offender Management Service.

Ioannidis, J. P. A. (2013). Clarifications on the application and interpretation of the test for excess significance and its extensions. *Journal of Mathematical Psychology*, *57*, 184-187. DOI: 10.1016/j.jmp.2013.13.002

Koehler, J. & Lösel, F. (2025). A meta-evaluative synthesis of the effects of custodial and community-based correctional treatment. *European Journal of Criminology*, *22*(1), 3-29. DOI: 10.1177/14773708241256501

Koehler, J. A., Lösel, F., Akoensi T. D. & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology*, *9*(1), 19–43. DOI: 10.1007/s11292-012-9159-7

Landenberger, N. A. & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, *1*(1), 451–476.

Lipsey, M. W., Landenberger, N. A. & Wilson S. J. (2007). *Effects of cognitive-behavioral programs for criminal offenders.* Campbell Systematic Reviews. DOI: 10.4073/csr.2007.6

Lösel, F. & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1*(1), 117–146. DOI: 10.1007/s11292-004-6466–7.

Lowenkamp, C. T., Flores, A. W., Holsinger, A. M., Makarios, M. D. & Latessa, E. J. (2010). Intensive supervision programs: Does program philosophy and the principles of effective intervention matter? *Journal of Criminal Justice*, *38*(4), 368-375. DOI: 10.1016/j.jcrimjus.2010.04.004

Lowenkamp, C. T., Latessa, E. J. & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime and Delinquency*, *52*, 77-93.

Lowenkamp, C. T., Latessa, E. J., & Smith, P. (2006). Does correctional program quality really matter: The impact of adhering to the principles of effective intervention. *Criminology and Public Policy, 5*, 201-220.

Mallion, J. S., Wood, J. L. & Mallion, A. (2020). Systematic review of 'Good Lives' assumptions and intentions. *Aggression and Violent Behavior*, *55*, DOI: 10.1016/j.avb.2020.101510

McGuire, J. (2013). "What Works" to reduce reoffending: 18 years on. In L. Craig, J. Dixon & T. A. Gannon (eds.), *What Works in Offender Rehabilitation: An Evidence Based Approach to Assessment and Treatment*. Pp. 20-49. Chichester: Wiley-Blackwell.

McGuire, J. (2025). Rehabilitation and processes of change. In L. Craig, J. Dixon & T. A. Gannon (eds.), *The Wiley Handbook of What Works in Correctional Rehabilitation: An Evidence Based Approach to Theory, Assessment and Treatment*. 2nd edition. Pp. 133-146. Chichester: Wiley-Blackwell.

Netto, N. R., Carter, J. M. & Bonell, C. (2014). A systematic review of interventions that adopt the 'Good Lives' approach to offender rehabilitation. *Journal of Offender Rehabilitation*, *53*(6), 403-432.

Petrosino, A. & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology, 1,* 435-50.

Polaschek, D. (2024). Eight criticisms of the recent review of the evidence base for RNR principles. *IACFP Newsletter*, August 23, 2024.

Schmucker, M. & Lösel, F. (2015). The effects of sexual offender treatment on recidivism: an international meta-analysis of sound quality evaluations. *Journal of Experimental Criminology*, *11*(4), 597-630. DOI: 10.1007/s11292-015-9241-z

Serie, C. M. B., Van Damme, L., Pleysier, S., De Ruiter, C, & Put, J. (2021). The relationship between primary human needs of the Good Lives Model (GLM) and subjective well-being in adolescents: A multi-level meta-analysis. *Aggression and Violent Behavior*, *61*, DOI: 10.1016/j.avb.2021.101651

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Weich, V., Kristjansson, E. & Henry, D.A. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *British Medical Journal*, 358. DOI: 10.1136/bmj.j4008

Stewart, L. A., Gabora, N., Kropp, P. R. & Lee, Z. (2014). Effectiveness of risk-needs-responsivity-based family violence programs with male offenders. *Journal of Family Violence*, *29*, 151-164. DOI: 10.1007/s10896-013-9575-0

Ward, T. & Stewart, C. (2003). Criminogenic needs and human needs: A theoretical model. *Psychology, Crime & Law*, *9*, 125-143. DOI:10.1080/1068316031000116247

Weisburd, D., Farrington, D. P & Gill. C. (eds.) (2016) *What Works in Crime Prevention and Rehabilitation: Lessons from Systematic Reviews*. New York, NY: Springer Science & Business Media.

Willis, G. M. & Ward, T. (2013). The Good Lives Model: Does it work? Preliminary evidence. In L. Craig, J. Dixon & T. A. Gannon (eds.), *What Works in Offender Rehabilitation: An Evidence Based Approach to Assessment and Treatment*. Pp. 305-317. Chichester: Wiley-Blackwell.

Willis, G. M. & Ward, T. (2025). Evidence for the Good Lives Model in supporting rehabilitation and desistance from offending. In L. A. Craig, J. Dixon & T. A. Gannon (eds.), *The Wiley Handbook of What Works in Correctional Rehabilitation: An Evidence Based Approach to Theory, Assessment and Treatment*. 2nd edition. Pp. 299-309. Chichester: Wiley-Blackwell.

Wilson, D. B. (2016). Correctional programs. In D. Weisburd, D. P. Farrington & C. Gill (eds)., *What Works in Crime Prevention and Rehabilitation: Lessons from Systematic Reviews*. Pp. 193-217. New York, NY: Springer Science & Business Media.

Wilson, D. B., Bouffard, L. A. & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior*, *32*, 172-204.

Wise, P. H. (2016). Cancer drugs, survival, and ethics. *British Medical Journal*, *355*, i5792. DOI: 10.1136/bmj.5792

Zeccola, J., Kelty, S. F. & Boer, D. (2021). Does the good lives model work? A systematic review of the recidivism evidence. *Journal of Forensic Practice*, *23*(3), 285-300.