|  |
| --- |
| **Concerns about the Methodology**  **AC:** The authors responded by changing the citation for the phrase "grounded, thematic approach" from Charmaz to Braun and Clark and explaining that the citations had been swapped by mistake.  Although this addresses one issue, it does not address the fact that a bottom-up thematic analysis—although commonly used in qualitative empirical research in CSCW—remains an unusual choice for a systematic review and **the authors did not provide examples of others who have done this**. It's also the case that the figures, counts, percentages, and cross-tabs look much more like a theory-driven or top-down content analysis than a Braun and Clark-style thematic analysis.  Of course, there is lots of guidance for conducting systematic reviews of this type. Chitu Okoli has published a very influential paper on conducting systematic reviews in information systems:  <https://aisel>.aisnet.org/cais/vol37/iss1/43/  Kitechenham et al. has published another on software engineering that has been used by HCI researchers:  <https://www>.sciencedirect.com/science/article/pii/S0950584908001390 An example of review following from HCI is Park and McKilligan: <https://doi>.org/10.1007/978-3-319-91797-9\_50  The authors’ approach isn’t clearly drawn from either of these methodologies, the original methodology from the Cochrane Library uses for biomedical research (where systematic reviews were pioneered), or elsewhere, as far as I can tell.  I understand that systematic reviews are still unusual in social computing research. I did find one recent review by Fitzpatrick and Ellingsen that used a thematic analysis but their approach (a) feels much more synthetic and interpretative (similar to the type of reviews published by Annual Reviews), and critically (b) does not claim to be a systematic review at all. https:/doi.org/10.1007/s10606-012-9168-0  With its counts, percentages, and figures, this paper feels like an Okoli-style systematic review. If that’s the goal, the review really should make it clear that ti has followed an established systematic review protocol and should more details on how this happened. If that’s not the goal, the there’s a mismatch somewhere.  Given the small amount of detail the authors provided on their methodology—unusual for systematic reviews—the fact that the authors did not provide examples of other systematic reviews that acted as a template struck me as an important missed opportunity.  **Problem with the Unit of Analysis**  **AC:** The issue I raised is about the unit of analysis. This paper's findings, figures, and interpretation are built on a series of quantitative counts and percentages of \*papers\*. As a result, a single system can be counted many times if the subject of many papers. There is some evidence that this might be happening. For example, you mention that 6 papers (12% of the sample) focused on CANS and 8 papers (14%) looked at SDM.  If 25 of your 50 papers focused on a single GLM-based system that ignored foster parent characteristics—perhaps all 25 papers were all written by the same productive research team)—you would correctly conclude that 50% of the \*papers\* fell into that those categories and the cross-tab. But if the system described was unique among systems in these features, the 50% number would really just reflect the fact that a single system with these features had been published more than other systems.  The problem is not small. Every single code and table is about a system and its methods, predictors, what outcomes. They are qualities of \*systems\* described in papers and not qualities of the papers. **Unlikely most systematic reviews, your conclusions seem to not be about gaps in the literature or what people should or should not be studying. They are about gaps in the kinds of systems being used in the foster care/CPS space.**  **How many distinct systems are described in your 50 papers?** If the paper had made it clear the 50 papers were about 50 distinct systems, I would withdraw my concern. That seems to not be the case and I'm still wondering how big this issue.  **Theory vs Empiricism**  **2AC**: the theory-vs-empiricism discussion and some related points don't sit quite right, especially with the description of caseworkers needing to make decisions that take into consideration additional information that isn't in the models (more empirical) at the same time that caseworkers are likely employing their own folk theories (complete with implicit biases) in their rationale for decisions. Do the caseworkers even have access to the underlying logic of the models, so that they could use that in making their own assessments? Or is it a black box that they have to just work around because its logic is opaque? In criminal justice algorithms, using lots of data with opaque algorithms has produced poor results, but simplified models that are easily understood by decision-makers seem to have good promise. In reality, both theory and empiricism are probably needed; theory-driven work might identify the additional data that could make models perform better, for example. **So casting this as a dichotomy seems like an oversimplification.**  **R1**: R2 pointed out that there are "doubtless better reasons for the criticism" than academic love of theory and the authors responded with changes in section 2.4. I didn't find these changes particularly convincing. They wrote "this is especially problematic in child welfare because each case of child neglect/abuse is contextually different and cannot be evaluated using the same set of significant predictors that are derived empirically [28]." I don't see how the existence of variation in CPS cases invalidates empirical derivation of predictors. I quickly read [28] and found a rich discussion of statistical issues in CPS risk assessment models, including the issue of omitted variable bias and a fascinating critique of how risk assessment models interact with institutionalized culture of risk-aversion in CPS. But I did not find support for critique of empirically derived predictors or statistical significance. More generally, I think the dichotomy between "empirical" and "theoretical" implicit in this critique is misguided, especially with "social science" on the side of theoretical. Useful theories will not be devoid of empirical content and people will not go through the effort to collect data without an a-priori notion (theory) that the data will be useful for a task.  **R1**: They addressed the statistical misstatements that I pointed out in my review. However, these were only meant to serve as examples of a problem with lack of clarity and precision in technical language throughout the paper. The paper still has some issues. It isn't clear whether the categories of "outcome variables" refer to dependent variables in the models or to target applications of systems. "Risk assessment" suggests that some negative outcome measure is a dependent variable in the model, but the other three categories do not seem like dependent variables to me but rather "outcomes" in a different sense as ways that systems are intended to be used. I had to check some references to see that at least with respect to the CANS tool, it seems like these are approaches that seek to optimize outcomes other than negative consequences. It might help if the authors can clearly point out specific measures that are used by the different models as examples of their categories of outcome variables. Similarly, as mentioned above, it isn't clear what is meant by "empirically derived" or "significant" predictors or how this relates to the need for theoretically motivated applications.  **Contribution to CSCW**  **AC**: The main contribution seems to be providing an introduction to CPS for CSCW researchers and points to a number of ways to go beyond this by drawing similarities, parallels, or difference to other social computing systems studied by CSCW.  **2AC**: It would be much more compelling if the authors could show how problems in CPS parallel other spaces in which CSCW work is being done, which is an opening for generalizable contributions across contexts, or how it's unique compared to those spaces. Similarly, if the problems in CPS were shown to have a clearly CSCW-relevant angle to them, that would also work -- as presented, one of the "hooks" is that caseworkers are being left out of the process of developing models. While this is certainly something that CSCW engages, it's also an example of the broader tech management issue of failing to include stakeholders into the conversation, which is by no means a new problem nor CSCW-specific.  Is there a case team component that I'm missing? I only saw a team discussed in the section on trauma cases, so if CPS is generally a single caseworker making decisions, where's the cooperative angle? There is literally no mention of collaboration or cooperation in the text -- this is really more of a decision support system, which isn't necessarily a CSCW thing. If multiple parties in case management were trying to use the algorithms in a team process, it would more clearly fall into the CSCW problem space.  **R1**: The authors did a good job deepening their engagement with CSCW literature, but I did not find much improvement to explaining the theoretical contribution to CSCW, as the authors think that a theoretical contribution is not necessary. Instead it aims to attract CSCW researchers to engage with the field of computational applications for CPS. In their summary of changes, the authors cite Wobbrock and Kientz to justify this claim, but immediately following the sentence they quote, Wobbrock and Kientz say "Survey contributions are appropriate after a topic has reached a certain level of maturity." This is fair, but points to what I'm coming to understand as a challenge in placing the contribution of this paper: it is an unusual survey because it surveys one field for an audience of a different field. A typical review synthesizes literature from one field for an audience of practitioners in that same field. So while the field of social work informatics might be relatively mature, the field of CSCW research into social work informatics is not. A well-executed survey of literature that is already of interest to a field is likely to attract interest, but a survey of foreign literature may not. So I am still concerned about how this paper will appeal "to members of CSCW who are not already interested in the foster care system."  **Minor issues**  AC: "Next, we define the inclusion criteria for the term "computational method" and the need for integrated data systems to conduct computational modeling." You're not defining the inclusion criteria \*for\* the term. I'm not sure what you mean.  AC: The phrase "this jives well with standard definitions" in §2.2 (p. 3) seemed very informal.  AC: Although the new graphs are a big improvement, they are also very big! And much of the space is empty.  AC: In the second to last sentence of §5.3 on pg 18, you have two citations inside a quotation. What's going on? Is this a quote from two places?  AC: describe caseworkers are producing "detailed ethnographic case notes." Really? Notes are \*not\* ethnographic just because they are unstructured text and captures contextual or subjective information.  R3: The manuscript now needs a thorough editing. Perhaps because so much revision took place in such a short period of time, there are words missing here and there, and some of the writing is rough. For example, section 5.3 could use a more general introduction. All of a sudden, we read a sentence beginning “Second,” without seeing a “First,” and without knowing what is being enumerated (beyond the information in the label). |