To: The Editors
From: John Wasson
Re: Response to requested changes

1. Administrative Clarifications:

Author Contribution:	Wasson –	Primary author and analysis of Medicaid and private practice validation sub-study
	Ho –	Writer and oversight of private practice validation data gathering
	Soloway –	Primary analysis of Medicaid data
	Moore –	Writer and oversight of Medicaid data gathering

Funder Contribution: "3 M Health Information Systems provided support in the form of salaries for authors [LS and LGM], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. This affiliation does not alter our adherence to PLOS One Policies on sharing data and materials."

Data Repository: The data relevant to the Medicaid and private practice analyses of risk for costly care is provided in www.HowsYourHealth.org/WhatMattersIndex.xlsx

2. Manuscript Editing: The document has been extensively edited in response to reviewer and editor comments. (See Attached revision).

3. Responses to Editors and Reviewers:

a) Please describe the number of patients who were invited to complete www.HowsYourHealth.org, and whether there are important differences in demographic, socioeconomic or clinical characteristics between those who completed the WMI and those who did not.

During 2014, Medicaid enrollees were sequentially approached and requested to complete the survey. Those included in the analysis were a subset whose responses to the survey indicated that they met entry criteria for chronic illness. Their socioeconomic and clinical characteristics are described in the text and in even greater detail in the Supporting Information. In actual practice, patients who complete assessments are selected as a convenient population sample, although they may not be representative of those who are either incapable of completing the survey, choose not to do so or do not use the practice.

Regression analyses were conducted to demonstrate that age, diagnostic burden, gender, and poverty level had little influence on the analyses of the outcomes from the Medicaid patients with chronic conditions. Nevertheless, to mitigate the possibility that this conveniently selected sample somehow biased the results, we provide additional comparisons of WMI performance in two very different populations: patients from private practices with chronic conditions and Medicaid patients with no chronic conditions. The results from these populations, which were similar to those from the primary analysis, are summarized in the discussion and detailed in the Supporting Information.

In Analysis:

"We do not know the characteristics of the patients who were not solicited or who were either unable or unmotivated to complete the assessment. Therefore, we used logistic regression to examine if the WMI's capacity to predict emergency or hospital use might have been vitiated by variations in the respondents' self-reported characteristics of age, gender, number of chronic conditions (listed above), and poverty (i.e., sometimes or always not able to pay for essentials such as food, clothing, or housing). To supplement this validation, we also examined the WMI's replicability in two very different patient groups: 1061 older patients from nine private practices and 4428 Medicaid patients without chronic conditions. The Supporting Information provides detailed information comparing the supplemental analyses with the primary analysis focused on Medicaid patients with chronic conditions."

In Discussion:

"Several limitations of the WMI deserve comment. First, these prospective results were derived from a self-selected sample of Medicaid patients with chronic conditions who completed a health assessment; whether the WMI would perform similarly for different patients is a valid concern. To address this point, we examined the WMI in 1061 older patients from nine private practices, selected using the same criteria for chronic conditions, as well as in 4087 patients with no chronic condition from the Midwestern statewide Medicaid program. Among the private practice patients with chronic conditions, the odds ratios (with 95% confidence intervals compared to a WMI of 0) for subsequent emergency room use were 1.8 (1.1–2.8), 2.1 (1.2–3.6), and 3.0 (1.4–6.3) for patients with WMIs of 1, 2, or ≥3, respectively. For WMIs of 1 or ≥ 2, the odds ratios of subsequent hospitalization were 1.4 (0.8–2.6) and 2.4 (1.2–4.5), respectively. In the Medicaid population without chronic conditions the odds ratios (with 95% confidence intervals compared to a WMI of 0) for subsequent emergency room use were 1.2 (1.03–1.40), 2.2 (1.73–2.76), and 3.2 (2.01–5.21) for patients with WMIs of 1, 2, or ≥3, respectively. For WMIs of 1 or ≥ 2, the odds ratios of subsequent hospitalization were 1.1 (0.87–1.48) and 1.6 (1.10–2.26), respectively. To summarize, the WMI's prospective prediction of costly usage was replicable in three very different populations. The Supporting Information further details the characteristics of all patients and the supplemental analysis procedures."

b) <u>Many papers describing risk models include a measure like the c-statistic that helps summarize the trade-offs between sensitivity and specificity. Please compute and report the c-statistic of each of the measures to help the reader compare the performance of these measures.</u>

The c-statistic is included in the revised paper to compare the overall accuracy of the proprietary CRM and the WMI for the Medicaid patients with chronic conditions. However, although the c-statistic is often assumed to indicate a model's clinical utility, in practice, modest c-statistic gains are offset by the considerable resources expended toward that objective. Therefore, we emphasize positive predictive value because this measure can be used to immediately and unambiguously inform appropriate decision-making.

In Analysis:

"In the Medicaid patients with chronic conditions, we compared the WMI and representative CRMs in three respects. First, we compared the models' sensitivities and positive predictive values for costly care. Sensitivity is defined as the proportion of patients who actually used costly care who were correctly designated as at-risk by the model. Positive predictive value is defined as the proportion of patients designated as at-risk who actually became users of costly care. Because predictive values are influenced by prevalence, we adjusted the WMI or CRM test thresholds so that the comparisons would be based on similar proportions of "at-risk" patients.

Second, we analyzed the distribution of the five WMI measures among patients designated by the CRMs to be at higher or lower risk. Third, we examined the relationship between true positives and false positives using the area under the receiver operation characteristic curve (AUROC), expressed as the concordance statistic (c-statistic). The c-statistic is frequently used to compare CRMs. The c-statistic approximates the overall accuracy of a binary classifier as its discrimination cut-point is varied. The c-statistic of a perfect classifier is 1.0; a c-statistic of 0.50 indicates classification is no better than chance."

In Results:

"The overall accuracies (c-statistics) of the proprietary CRM and the WMI were the same (0.63). (The c-statistic cannot be calculated for Medicare CRM because its cut-point is fixed)."

c) <u>The paper would be strengthened by including analysis of a hybrid model including both a traditional CRM and the WMI score. As it stands, we cannot see if WMI captures the same variation in outcome risk as CRMs, or if there is some non-overlap. A hybrid model would still allow the benefits of WMI (e.g. allowing patients to communicate barriers/needs to care teams), and could potentially maximize the performance of the model to identify patients who need additional help.</u>

The distribution of WMI measures among CRM risk groups presented in Figure Two addresses the reviewer's interest in "overlap" in a clinically meaningful way.

In addition, previously published hybrid models have generally been based on pre-existing data because gathering, integrating, weighting, and clinically applying the information prospectively is extremely difficult and laborious. In our example, there are three ways a hybrid model might be applied in a clinical setting: WMI and CRM analyses for all patients; CRM for all patients followed by WMI for only those patients designated as at-risk by the CRM; or WMI for all patients followed by CRM for only those patients designated as at-risk by the WMI. Adding to this administrative complexity, the Data Repository shows the 40 different categories of measures (4 WMI and 10 deciles of the CRM) that would require combination in a “black box” artificial intelligence program to make predictions ($4 \times 3 \times 2 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2$ possible combinations).

Therefore, any hybrid model must achieve a very high rate of accuracy to compensate for this level of complexity. The literature reports that hybrid models may increase the c-statistic marginally, but the resultant positive predictive values remain clinically weak and do not justify the resources required to achieve the marginal gain.

The following text has been added in this revision to address the overlap between the methods and the possibility of hybridization.

In Results:

“When a CRM based on clinical and laboratory data is combined with patient-reported data, an increase in the c-statistic is often documented (7,8,27). However, in practice, incorrect classification persists even with such hybrid models, and the c-statistic gain is offset by the considerable effort required to combine the clinical, laboratory, and patient-reported inputs such that the output can be made available in a manner that is timely and useful for clinical practice. For example, when a hybrid risk model is established by combining a WMI ≥ 3 with the proprietary CRM’s highest risk decile, 290 (67%) and 186 (43%) of the 431 patients thus classified as the highest risk subgroup subsequently used emergency or hospital services, respectively. These predictive values represent only small improvement over those of the CRM highest risk-decile alone (59% and 37%, respectively). ”

d) <u>Especially, the authors need to elaborate more regarding implications based on their findings. Please put into context the performance of the WMI and the CRM measures by offering some comparison to other published measures that have been used to assess risk for hospitalization or ED utilization.</u>

We have added a new section of text in the discussion that contextualizes the WMI within the tradition of self-reported, case-finding instruments that have been applied to certain specific populations. The “Illustration” section further contextualizes the WMI as it has actually been applied in clinical settings, in comparison to the examples presented in other sections of the paper that illustrate the shortcomings of CRMs not based on self-reporting.

Discussion:

An illustration:

"Each WMI item is meant to elicit an action that meets each patient's needs. A very common and remediable risk factor included in the WMI is a patient's lack of confidence in their ability to manage most health concerns; this risk factor is associated with many adverse health experiences, including more frequent (and often avoidable) emergency or hospital care use, lost time from work, and medical harm [13]. Applying the WMI model, patients who say they are not confident that they can control and manage most of their health problems are then asked by medical assistants or the online health assessment (www.HowsYourHealth.org) to answer the query, "What would it take for you to be able to say that you are very confident that you can control most of your health problems during the next two months" [12,20]? Their verbatim responses are included in a summary report for the clinicians who provide their care. Examples of queries for the other WMI items are listed elsewhere [9].

To illustrate how the WMI identifies population needs in a clinical setting, we summarized the verbatim responses to the online assessment of 1915 adult patients from across the United States. These patients met the identical selection criteria that was used to select the Medicaid population sample. The verbatim responses, in which patients identified the health care interventions that they perceived would be most effective, could be generally classified into the following four categories.

1) *Professional help. Patients most often request better medical information and education: clarification of their diagnoses, timely sharing of test results, and when possible, additional relief of symptoms. Examples: (a) "Help of a doctor who will actually listen and take my problems seriously without just pushing medication." Michigan; WMI = 2. (b) "If I got an accurately diagnose of my illness, and able to get a specific course of treatment I could control and manage my health problems." Texas; WMI = 2.*
2) *Personal change. Patients acknowledge their need to improve time-management, motivation, and lifestyle. Examples: (a) "Staying focused on what is required to be healthier." New Hampshire; WMI = 2. (b) "More time and attention to my diabetes." North Carolina. WMI = 2.*
3) *Non-professional support and guidance. Patients request coaching or support in the workplace, home, and/or community; financial assistance may also be needed. Examples: (a) "Finances are stopping me from getting medical help. Co-pays for doctors and medications has taken most of my life savings." Rhode Island; WMI = 2. (b) "Need some coaching." Minnesota; WMI = 2.*
4) *Non-response or uninterpretable response.*

Fig 3 compiles 1915 patients' verbatim responses regarding changes they require to improve their health confidence, and illustrates how their needs vary in relation to their WMI sums.

For this sample population of adult patients with chronic conditions, higher WMI sums are strongly associated with an increased likelihood that the respondents identify a need for professional assistance, and with a reduced likelihood that they consider their personal behavior as the primary remediable cause for their low confidence. Logistic regression confirms the persistence of this pattern ($p < 0.001$) regardless of patient age, gender, financial status, or number of chronic conditions. For patients who used hospital or emergency care in the past year *and had a WMI ≥ 3, half (70/142) believed that the event may have been avoidable; for those having a WMI = 1, approximately one in five (15/76) shared that belief.*

In summary, people simply answer each question and bring their responses to the attention of someone who can help them address each problem, such as a health professional, a support group, a knowledgeable friend, or even a website like the one used to develop the WMI (HowsYourHealth.org). Services appropriate for the level of risk based on "what matters" is the goal. Thus, as this illustration demonstrates, to improve low health confidence more attention to medical diagnostics, therapeutics, and education is indicated when the WMI is high; more support for behavior change when the WMI is low. (An example of the WMI for public distribution is included in the Supporting Information)"