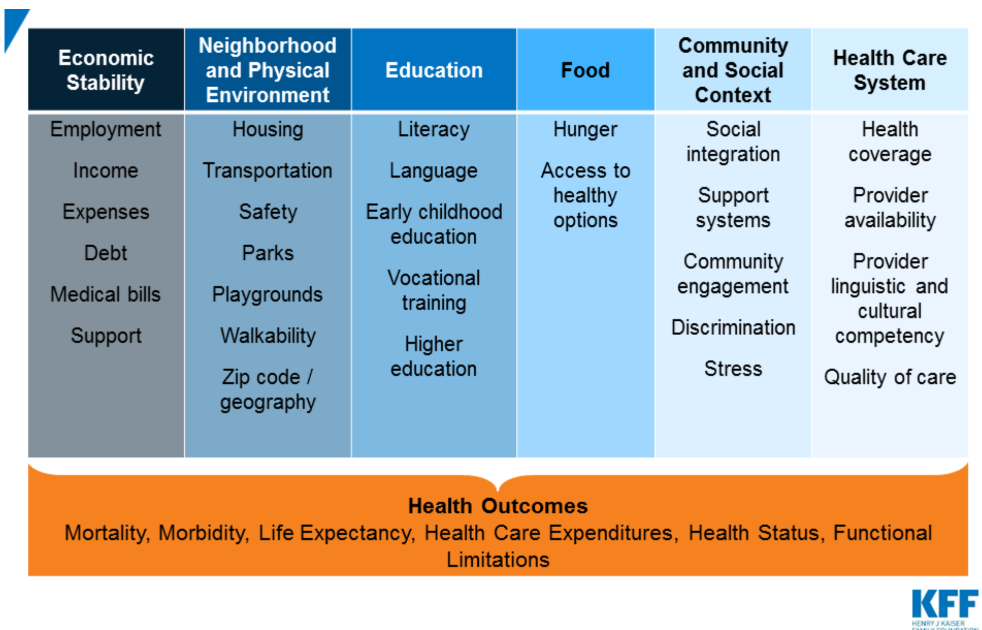
The 2024 Community Companion Challenge: Transforming Care with Data-Driven Compassion

Background

Social determinants of health (SDOH), such as where we live, income, education, race, age, and support systems have a profound impact on our health and well-being. These factors contribute to health disparities and are strong predictors of both individual and community-level health outcomes. Understanding an individual's SDOH is crucial, especially when providing care for those with disabilities, chronic conditions, or when supporting aging loved ones. However, obtaining a comprehensive picture of someone's social circumstances can be challenging. Current approaches often rely on surveys or self-reporting, which have limitations and can be time-consuming.



Why are SDOH important?

- Health Disparities: Addressing SDOH is essential for reducing health disparities and achieving health equity. It recognizes that health is not just a product of genetics or personal choices but is deeply influenced by social and economic factors.
- Overall Health: SDOH influence chronic disease development, treatment outcomes, mental health, and well-being.

- Healthcare Costs: Unmet social needs drive increased healthcare utilization, emergency visits, and preventable readmissions.
- Preventive Approach: Understanding SDOH allows for a more preventive approach to health, targeting the root causes of poor health outcomes rather than just treating diseases as they occur.

How are SDOH currently used?

- Patient Navigators (PNs) use standardized inventories like PRAPARE-English.pdf, Health Leads Screening_Toolkit_2018.pdf, and AHCM-HealthSocialNeedsScreeningTool_1_8_18.pdf to capture a patient's own answers about their social determinants. Alternatively, they may ask the users to self-report.
- PNs or physicians may then look at any flags and coordinate with internal team members such as a social worker (if they have one) or another organization to talk more about that gap and coordinate referrals. Most often, physician offices just give a list of nearby agencies to contact and that's that.
- Some physicians customize interventions based on social needs identified through screening (e.g., if an individual lives in a rural area, providers may prescribe a 60 day medication supply instead of 30 days).

Challenges faced with using SDOH at present

- **Feasibility Issues:** Time constraints and insufficient reimbursement make it challenging for patient navigators (PNs) and physicians to administer SDOH surveys effectively.
- **Patient Reluctance:** A widespread hesitance to share personal information, alongside communication barriers and varying navigator expertise, limits effective screening.
- **Binary Evaluation Limitations:** Common SDOH inventories (e.g., PRAPARE, Health Leads) use binary evaluations that fail to capture the nuanced differences within variables, overlooking the varied risk levels among different groups. For example, within race, risk differences exist between whites -> asians -> mixed race -> others.
- **Ignoring Interdependencies:** Existing tools evaluate SDOH factors in isolation, missing the interconnectedness of issues like transportation, work commute, and income, which could lead to inaccurate intervention strategies. For example, transportation challenges may be significantly worsened with accompanying low income and long commute.
- **Actionable Guidance Deficit:** Healthcare providers often lack clear, actionable steps and resources to address identified social determinants or risk factors in patients. How

do these social parameters influence their clinical decisions such as tackling appointment no shows, medication compliance, or patient education?

- **Lack of Self-Service Tools:** There's a significant gap in empowering patients and families to navigate SDOH challenges independently, such as in finding transportation solutions beyond generic tools like Google Maps.

Example scenarios of SDOH in action

Scenario 1: Jamie's Transportation Barriers

Jamie, a single mother juggling multiple jobs, knows she needs that follow-up appointment. But with no car, patchy bus service, and childcare costs that eat up most of her paycheck, it feels impossible. A predictive model for social risks can flag transportation risk. Here's where we go beyond the obvious and help the physicians in making an informed choice. A physician may be debating about the right option to support Jamie, such as -

- Childcare: More than just the ride, it's partnering with community programs for on-site childcare during appointments. Is it likely that Jamie needs childcare support?
- Beyond the Office: Could a mobile clinic meet Jamie in her neighborhood? This addresses transportation and potentially offers more flexible hours.
- Efficiency AND Agency: Consolidate appointments, offer longer medication supplies, BUT let Jamie choose telehealth when it works for her schedule. However, is Jamie likely to have internet or cell phone access?
- Transportation: Should we provide transportation vouchers to Jamie? Does her insurance cover it?

Real-life impact: For example, if we determine the last one is likely a better option based on higher risk for a transportation gap, we can proactively share nearby transportation options, summarize information from insurance benefits (if we have them), and also give her a voucher (if she is eligible).

Why it's powerful: We highlight the intersection of needs (transportation, childcare, work) and that solutions respect the individual's situation.

Scenario 2: Martha's Post-Discharge Need

Picture Martha, a frail and overwhelmed 70 year old senior after being discharged from a hospital. Nutrition is very important after a hospital discharge and Martha may be eligible for free meal delivery from a nearby non-profit organization, such as Meals on Wheels or other food bank. But physicians don’t have an easy way to predict if Martha may be eligible for such services in her neighborhood (e.g., low income, certain age threshold, veteran status, etc.).

Each non-profit may have their own criteria. Moreover, patients and families don’t know which non-profit may be accepting new patients for food delivery (not everyone can because of resource constraints). The delays in assessing eligibility, finding an organization nearby that is accepting new applicants, and completing paperwork often leads patients to be without any support for a couple of weeks. Our SDOH prediction can change this:

- Risk to Action: Based on her address we determine need, which is further solidified by the risk model once we add additional demographics, such as age, income (if we have the data). We don't wait, preemptively cross-checking benefits to streamline the process.
- Beyond Meals: Are there local food pantries for shorter-term needs during the wait? Is there a senior center with communal meals, addressing both nutrition and isolation?
- Proactive AND Respectful: We inform Martha and her doctor, AND send a text or notification to shortlisted food service organizations to facilitate a quick decision.

Real-life impact: In this situation, we can share benefits information with Martha and point her to Meals on Wheels or other nearby food services, while also guiding the physicians AND the organizations about her eligibility.

Why it's powerful: We show how quick action matters, and that "support" has layers Martha might not be aware of.

Scenario 3: Arthur's Unseen Risk

Arthur, homebound due to his disability, is easy for the system to overlook. Arthur may find it very inconvenient to frequently visit the physician’s office and unless symptoms appear, he may not feel the need to go in-person. However, using our social scorecard, we determine that Arthur lives in a neighborhood, which has a much higher incidence of high cholesterol. Instead of waiting for Arthur to face a complication, we can proactively flag this potential risk.

- Meeting Him Where He's At: Can in-home blood draws be arranged? His address can help find services that offer them.
- Prevention as Empowerment: Information on heart-healthy living tailored for someone less mobile, NOT just a scolding to "go see a doctor."
- Bridging the Gap: If a doctor visit IS needed, are there providers specializing in serving patients with disabilities? Proactive match-making is key.

Why it's powerful: We highlight how SDOH data prevents problems before they become urgent, and that care extends beyond the clinic walls.

Real-life impact: In Arthur’s case we can select and share information about the top 3 endocrinologists near him as well as a couple of transportation options. We can also share basic information about cholesterol screening tests and help Arthur schedule a lab test.

Challenge Details

Our quest is to use the power of data science to provide a new way of understanding social determinants without relying solely on asking the individual about their needs. Such a proactive model can not only save time for the physician staff, but also enable direct support and guidance for the patients and their family members.

Objectives

1. **The Heart of the Challenge - The Social Care Scorecard:** Imagine a tool that uses data to paint a picture of an individual's social support needs and potential health risks.
    a. Predictive model for a social scorecard based on limited available information about an individual and flag needs like transportation, healthy food options, etc.
    b. An action-based version of the scorecard to assist PNs, physicians, and other clinical staff (based on their role) in taking action.
2. **Making a Real-World Difference - Connecting People with Resources:** The social care scorecard isn't just about insight; it's about action. You will pool relevant information and resources from provided sources. This will ensure the scorecard helps identify the next best course of action and match people to essential support services.
3. **The India Connection - Data for a Healthier Future:** India faces unique healthcare challenges, where social factors play a crucial role. Brainstorm innovative data sources that can help power a similar predictive system for India to make a lasting impact.

The Heart of the Challenge - The Social Care Scorecard

Before starting the exercise, take some time to get familiar with the social determinants - What are they? How do they influence healthcare decisions?

Sources: *SDOH questionnaires in reference files; raw data sources outlined in Appendix 1; Patient test demographic data (see last bullet in Tasks below)*

Curating data for analysis

In order to design an algorithm, you may need to clean up the data and split it into at least two or three different segments - **training, tuning, and testing.** Here are some steps (indicative) you can take to create these datasets for your team -

1. Normalize the data based on the initial superset of parameters you want to analyze (maybe a larger number of parameters than your final result). This results in a clean superset of data.

2. Use clustering techniques to group data from certain states into clusters based on most important parameters that influence social determinants. This helps create datasets with internally similar states, allowing you to compare how algorithm performance might vary across different social determinant contexts. Examples of important parameters are -
    a. Socioeconomic - income, poverty level, educational attainment
    b. Demographic - race, age, urban / rural split, total population
    c. Health outcomes - life expectancy, chronic diseases, readmissions, hospital use
    d. Geographic diversity - North, South, West, Northeast
    e. Others - political affiliation (Democrat vs Republican; as there are significant differences in insurance), % of Medicaid (low income insurance population; can be a substitute for poverty), % of Medicare (substitute for 65+ population)
3. For sake of simplicity, you can choose the top 5 metrics to use to split the master dataset. This will make it more manageable *(be prepared to share your rationale).* No splitting method can guarantee perfectly comparable datasets due to the complex nature of social determinants. It's essential to be aware of the inherent variation. You may need to iterate a bit. This step may also help you identify the most important social determinants for further analysis.

Tasks

- Understand the parameters included in questionnaires such as PRAPARE or Health Leads that are often used to ask the questions to the individuals directly. Learn about the parameters that are most important in determining social risk.

- Draft an initial scorecard with your parameters of focus and the indicators that will help you get there. Ask / research these questions - Which parameters have a strong link to healthcare outcomes? Or which factors influence the physician's decisions about a person's care? Which factors influence the overall well being of a person to help them lead a long, productive, life? *Refer to some additional guidance in Appendix 3.*

- Next, select the most important parameters from provided data sources and normalize the data. ALL teams should use data pertaining to ALL US areas for this step, since there could be regional variations which may skew results if you focus on only one state or region. *You can also create a separate group-only version of the model, which may be more accurate at a regional level as well.*

- Assume you may know the following about an individual in this order. For example, we know about Jamie from 30319 at a minimum. (Jamie can be a guy, woman, non-binary so it is difficult to judge gender).

    a. Zip code (at a minimum). *The rest of the information may likely not be available, but the model should 'adapt' to such new information if it becomes available.*
    b. Address (maybe; need to map it against census tract and zip code)
    c. Gender

d. Age
e. Race
f. Income
g. Education
h. Veteran status (yes / no)

- ***Even though zip code data is available, it is strongly advisable to use census tract data for training and then map it as it has higher fidelity. Refer to Appendix 1.***

- **Generate patient test data:** Generate patient personas with Synthea database or use data from other research. You can also structure a strong prompt to generate example patient SDOH and demographic data with LLMs or multi-agent frameworks using GPT-4 and Gemini Advanced. These personas will help you test different scenarios and also demo output. DO NOT use this data to overfit on clinical parameters.

Output

A. **Predictive Modeling at the Census Tract Level:** Using user information (e.g., Jamie from zip code 30319), this model generates a preliminary risk profile for various social determinants at both national and state levels. It offers a composite score that breaks down into assessments of various sub-parameters. For instance, the evaluation of transportation risk might include factors such as vehicle ownership, commute times exceeding 30 minutes, and proximity to hospitals and pharmacies.

B. **Core Social Scorecard:** The aim is to craft a detailed scorecard, enhancing the general approach seen in tools like PRAPARE or Health Leads with greater specificity.

C. **Role-based Social Action Scorecard:** Building upon the base scorecard, this creates an actionable view for the clinical staff (aka answering 'so what'?). The intent is to help them make decisions about how they should plan appointments or share education, etc. *Refer to Appendix 3.*

D. *(Optional Stretch Goal)* **Social Trends and Future Needs Forecasting:** Analyze trends in social determinants for a particular census tract / neighborhood to flag any future high risk needs for the entire population living there (based on worsening patterns). This could highlight emerging issues, such as an increasing elderly population or diminishing access to transportation, enabling proactive community and healthcare planning.

Hint: Outline your scorecard before identifying parameters of interest to include in your analysis.

Making a Real-World Difference - Connecting People with Resources

For this next step, you will be assigned to a group. Your team will focus on building a database of community support resources and information for your given region. Please see the separate file for more information about the data sources.

Sources: See Appendix 2

Tasks

- You will be assigned one of these groups for this part of the challenge.
    a. Southern California (2 teams)
    b. Northern California, Hawaii
    c. New Mexico, Arizona, Nevada
    d. East & South Texas
    e. North & West Texas, Oklahoma
    f. Washington, Oregon, Alaska, Montana, Idaho
    g. Colorado, Utah, Wyoming, Minnesota, South Dakota, North Dakota
    h. Central & South Florida, Puerto Rico
    i. North Florida, Georgia
    j. New York (2 teams)
    k. Massachusetts, New Hampshire, Maine, Vermont, Rhode Island
    l. New Jersey, Connecticut, Delaware
    m. Maryland, Washington DC, Virginia
    n. North Carolina, South Carolina
    o. Pennsylvania, West Virginia
    p. Michigan, Ohio
    q. Illinois, Wisconsin
    r. Indiana, Tennessee, Kentucky
    s. Kansas, Missouri, Iowa, Nebraska
    t. Alabama, Louisiana, Mississippi, Arkansas

- Create a database for your group's region of focus using diverse data sources provided in Appendix 2.

- Once you have created a support database, you can use your predictive algorithm to include suggestions about nearby transportation options, physicians, home health agencies, etc. These suggestions will be based on an individual's scorecard, so they will ONLY get support suggestions IF they are at a Medium or High risk for that particular need.

- If you find any additional data sources that you feel are relevant, please just make a note of them in your presentation, but do not include them in your output.

Output

A. Include a section in your scorecard that highlights up to 3 suggestions from your support database based on each identified social needs. For example, nearby transportation providers with rating and contact OR nearest pharmacies. *You can take the center of the zip code for mapping the distance.*

B. *(Optional Stretch Goal)* Visualize your scorecard and matched findings.

    a. Design two interfaces to share the same findings and recommendations - 1) Patients and families and 2) Physicians and organizations.

    b. You can use generative AI agents and/or a react dashboard. This may also allow you to demonstrate various scenarios, where you enter different information about a user and showcase its impact on their social scorecard generated by your predictive model.

Final Presentation + The India Connection - Data for a Healthier Future

As you finalize your algorithm, scorecard, and mapping, we would love to know more about your team's approach to this challenge **in no more than 10-15 slides.**

Tasks

- Create a presentation that outlines your key assumptions, reasons for selecting or prioritizing certain parameters, and your approach to the problem. Also include analytical techniques you tested and why you selected a particular one.
- Outline some examples of your model and scorecard in action.
- Add your team's perspectives on what other data would you have added, if available to improve the predictive and personalization.
- Add your team's perspectives on the relevance of this exercise in the Indian context - What sources of data can provide required info or help estimate it? How can we use these determinants to create a more 'proactive' care approach? Can this also help with planning and allocation of resources towards building healthier villages, towns, and cities? *(2 to 3 slides)*

Output

A. Final presentation and findings along with a walkthrough of your solution

B. Accompanying document highlighting each segment of your code

Evaluation Criteria

This challenge is expected to yield innovative solutions that will significantly advance the field of community health, making personalized, proactive health support accessible to all. Here are the key evaluation criteria along with their respective weights in the evaluation.

1. **Normalization and design:** Selecting the right parameters and normalizing them is a very important step. We will be evaluating the soundness of your approach, including the options you considered and decisions made. (20%)
2. **Accuracy and predictive power of the model:** We will use a test sample of users to evaluate how accurately the model builds your scorecard. (25%)
3. **Quality of scorecard and matching:** As the focus is on ‘action’, how you design the scorecard and then use it to make recommendations on relevant community resources will be a crucial consideration. (25%)
4. **Comprehensiveness of the resource database:** The extent and relevance of the resources included in the database. (10%)
5. **Uniqueness of ideas:** Quality of your team’s ideas along the entire challenge and their uniqueness will be considered (incl. relevance to India). (10%)
6. **Demo and presentation:** Finally, how you succinctly explain your team’s work, present findings, and showcase your model’s abilities will be considered. (10%)

Final words

The Community Companion Challenge isn't just about code and algorithms. It's about empowering patients and families to navigate their health challenges with a little less stress and a lot more support. Your work has the potential to improve lives, reduce healthcare disparities, and create a more compassionate, data-driven healthcare system.

Are you ready for a challenge that makes a difference? 🙂

Appendix 1: Data Sources For The Heart of the Challenge - The Social Care Scorecard

Using codebook

- The codebook available here contains crucial information regarding variables in different sources below.
- **The 2 tabs 'All SDOH Variables' and 'All Health Data Variables' are your primary source of information.**
- All SDOH Variables: Col. B indicates the relevant social determinant, while Col. E indicates the variables in the dataset. NOT all variables are important for your analysis, and this is an evaluation criteria. Col. F provides some hints on relevance. In Col. E,
  - Light orange cells are very likely to be relevant in some way. You may need to compute a blended variable in many cases.
  - Other variables could be relevant based on your approach.
  - Light gray fonts / cells are deemed very less relevant. But in case you choose to select any of these parameters, please share your rationale.
- All Health Data Variables: Col. B indicates various variables, while Col. C includes some notes on relevance for different reasons in a clinician's workflow. These notes are NOT exhaustive.

Source 1 - Social Determinants of Health (SDOH) AHRQ data - PRIMARY

https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html

- Latest data is from 2020 - MOST IMPORTANT. 2019 and 2018 data has also been provided in case you want to consider trends for high impact social parameters.

Working with it

- Metrics based on Census Tract are preferred as they are likely to be most accurate when we have a user's address. And here is more information about Census Tract vs. Zip Code. Additional information is available here and here. Some metrics may only be available at County level (low fidelity), but you can consider them if important.
- Census tracts may need to be mapped to zip codes (as we won't know a user's census tract by default) using the included HUD 2020 Zip Code - Census Track Crosswalk.xlsx . Here is additional documentation about the crosswalk.
- This is the base data to create a baseline predictive algorithm for social determinants if we know a user's address. Assume that we will at least have an individual's zip code. For some we may have a full address. And then for some fewer we may have age,

gender, race, veteran status. This algorithm should then improve its predictions if we have more user details such as their age, race, or gender.
- You can consider using health outcomes data to determine which parameters have a positive or negative correlation with indicators for overall health and well-being, such as life expectancy (Source 3) and some of the others provided in Source 2.

Source 2 - Health Related Census Tract Data

https://healthdata.gov/dataset/PLACES-Local-Data-for-Better-Health-Census-Tract-D/jpdw-4rwm/about_data

- Latest 2023 release built on primarily data up to 2020.
- Based on Census Tract
- Includes data about disabilities, health outcomes, insurance, etc.

Working with it
- Importance: This is an add-on data that can further refine the baseline predictive algorithm and predict 'risk' of certain healthcare needs along with social determinants

Source 3 - Life expectancy by census tract

U.S. Small-area Life Expectancy Estimates Project – USALEEP - https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html

- Although this data is slightly older than the rest of the files, you can use it for adjusting your predictive model

Appendix 2: Data Sources For Making a Real-World Difference - Connecting People with Resources

Tips:

- *Focus on relevant information for your region.*
- *Even within your region, you can select a sample of zip codes that covers a large portion of the region (20-25 usually should be sufficient). You DO NOT have to look at every zip code in a region.*

Mapping user's zip code or address to a census tract

You can either create your own mapping from the provided HUD data (above) and/or also use APIs to map in real-time, such as those provided by US Census bureau - https://geocoding.geo.census.gov/geocoder/Geocoding_Services_API.html

Table of sources of data for the real-world mapping

Data Type	**Source**	**Notes**
Hospitals	Copy of Hospital_Gener…	Especially important parameters include general hospital details, address, contact, quality rating, hospital type. Additional info, if needed - https://data.cms.gov/provider-data/dataset/xubh-q36u
Home Health	Copy of HH_Provider_Ja…	You can use data like location, contact, types of services offered, star rating, discharge to community rate, potential readmission rate (key indicator of quality). Additional info, if needed - https://data.cms.gov/provider-data/dataset/6jpm-sxkc
Long Term Care	Copy of Long-Term_Car…	In addition to the name of facility, address, contact, etc. you can also use quality parameters like readmissions, discharge to community, and Medicare claims (costs). Additional documentation available here - https://data.cms.gov/provider-data/dataset/fp6g-2gsn
Assisted Living	https://www.caring.com/senior-living/assisted-living/	You can use the provided web scraper script to pull relevant information for your group. *Modify, if*

		needed.
Memory Care	https://www.caring.com/senior-living/memory-care-facilities/	You can use the provided web scraper script to pull relevant information for your group. *Modify, if needed.*
Independent Living	https://www.caring.com/senior-living/independent-living/	You can use the provided web scraper script to pull relevant information for your group. *Modify, if needed.*
Providers	https://data.cms.gov/provider-data/dataset/mj5m-pzi6	Download from the link for your region. You can use name, address, specialty, group practice, and other data as well as quality, telehealth, and whether they accept Medicare.
Suppliers	Copy of Medical-Equipm…	You can use name, address, contact, and also type of supplier or even example supplies offered when mapping to appropriate results (later data is unstructured, so consider using gen AI). Additional documentation available here - https://data.cms.gov/provider-data/dataset/ct36-nrcq
Food Services	https://www.mealsonwheelsamerica.org/	Use a provided web scraper to pull data for your group
Transportation	https://www.communityresourcefinder.org/	Use a provided web scraper to pull data for your group (under Community Services)
Adult Day Care	https://www.communityresourcefinder.org/	Use a provided web scraper to pull data for your group (under Community Services)
Elder Law Attorneys	https://www.communityresourcefinder.org/	Use a provided web scraper to pull data for your group (under Community Services)
Geriatric Care Managers	https://www.communityresourcefinder.org/	Use a provided web scraper to pull data for your group (under Community Services)
Area Agency on Aging	https://www.communityresourcefinder.org/	Use a provided web scraper to pull data for your group (under Community Services)
Home Care	https://www.communityresourcefinder.org/	Use a provided web scraper to pull data for your group (under Care At Home). These are different from Home Health and focus on general day-to-day support for seniors and disabled.

Appendix 3: Examples / Indicative Structures for Scorecard

You can design your scorecard in two ways -

1) To closely model after the prioritized structure of PRAPARE or Health Leads with composite traffic light view
    a) Such a scorecard may have categories like Economic instability, Transportation issues, Food insecurity, etc. along with their sub-parameters influencing risk (e.g., transportation)
    b) And it should have the healthcare related categories and recommendations (e.g., screening)

2) To focus on an issue-based approach where some of the categories are geared towards physicians to help them make decisions. As we discussed earlier in this document, existing PRAPARE and other documents fail to create a composite issue driven view for the providers to truly help them with decision making. This view changes that -
    a) For example, such a scorecard may have additional categories such as Appointment no shows (along with its composite risk + sub-parameter level risk - *see example below)*

Example of Appointment No Shows

Appointment no shows are very important as when the patient fails to show up, it leads to loss of time and revenue for the physicians. There may be a roll-up risk score for 'Appointment no show' (high) followed by a composite risk for 'transportation challenges', 'financial strain', and 'education challenges'

Determinants	**Example Indicators**	**Solution**
Transportation challenges	No vehicle in household Distance from nearest bus stop Workplace is >30 min away by public transportation (less so by own car) Driving distance from location to nearest hospital (less important) Actual driving distance or public transportation from user's address to the nearest hospital (using Google Maps)	Transport vouchers Offer tele-health (if connectivity is good)

Financial strain	Household income (based on avg. household size. E.g., <$29k for individuals or <$49k for a family of 3 is high risk) Distance from work to nearest hospital	Combine multiple appointments or lab tests Educate about financial aid Educate about long-term healthcare costs of complications if missed appointment Schedule appointments in the evenings or weekends
Education challenges	High school or less (high)	Simplified instructions & reminders Prevention & importance of various lab tests or follow-up visits
Technical access and risk	Access to phone data or internet Quality of internet in the neighborhood	If good, providers can offer virtual visits as an alternative (in case of high transportation risk).