Towards Efficient Interaction for Personal Health Data Queries on Smartwatches

Bradley Rey
bradley.rey@ubc.ca
University of British Columbia
Kelowna, Canada

Charles-Olivier Dufresne-Camaro
chcamaro@student.ubc.ca
University of British Columbia
Kelowna, Canada

Pourang Irani
pourang.irani@ubc.ca
University of British Columbia
Kelowna, Canada

ABSTRACT

The smartwatch is rapidly becoming a go-to personal health tracking device, allowing for the collection of a broad range of personal health data. Yet, access to this data is often limited to discrete glanceable visualizations. This in part is due to a lack in our understanding of the queries desired to access such data. Thus, as practitioners and application designers, our ability to enable efficient exploratory interactions is limited. In this work, through analysis of a public dataset, we characterize personal health data queries desired for exploration on the smartwatch across multiple dimensions: (i) data requested and attributes of this data, (ii) aggregation methods, (iii) mechanisms for filtering, and (iv) interrogatives used. We conclude with discussion around our findings that can be utilized in future works aimed toward enabling efficient interaction with personal health data on the smartwatch.

CCS CONCEPTS

• **Human-centered computing → Interaction design**; **Natural language interfaces**; **Interaction devices**.

KEYWORDS

smartwatch, interaction, natural language, personal health data

ACM Reference Format:
Bradley Rey, Charles-Olivier Dufresne-Camaro, and Pourang Irani. 2023. Towards Efficient Interaction for Personal Health Data Queries on Smartwatches. In *25th International Conference on Mobile Human-Computer Interaction (MobileHCI '23 Companion), September 26–29, 2023, Athens, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3565066.3608700

1 INTRODUCTION

As smartwatches become increasingly capable devices with regard to the collection of personal health data, interaction with said data directly on the smartwatch lags behind. While a smartwatch may not need to handle all exploratory tasks possible, there exists a broad array of exploratory needs which are desired by people to support unique health outcomes [1, 23, 29, 32]. Smartwatches, through advancing input modalities such as touch, speech, gesturing, and buttons/dials, have the potential to enable broader interaction with the collected data. Fundamentally, however, our lack of knowledge surrounding the interactive requirements for personal health data queries hinders progress.

Previous work has recognized the benefit that smartwatches carry during in-situ and on-the-go usage scenarios for actionable insight (e.g., while at the gym or on a hike) [1, 16, 22, 30, 32]. In fact, in these scenarios, smartwatch use can eclipse that of other devices which are better suited for interaction with data [11]. As such, to better enable smartwatch data exploration in these scenarios, research focuses on components and characteristics of glanceable and micro visualizations [1, 4, 7, 17, 27]. Additionally, singular tasks (i.e., max/min or single value detection) have been studied [28]. While beneficial in their own right, these works still remain limited in their interactive capabilities and access to broader insight. With research showcasing promise for multi-modal interaction with personal health data [21], and the affordance of these input modalities on the smartwatch, we focus on first better understanding desired personal health data queries to then be able to appropriately accommodate exploration through these interaction modalities.

Rather surprisingly, there are a limited number of datasets which capture desired personal health data queries [31, 32]. Often, the focus on analysis when looking at personal health data exploration needs is to uncover overarching themes and categories [1, 8, 9, 32]. These works are immensely impactful in focusing on *what* is desired for exploration. Furthermore, research has directed attention to learning outcomes, data collection practices, and issues encountered [10]. Yet, these works do not provide us the means to better understand *how* people naturally wish to query their data. This is crucial towards the goal of providing practical and expanded exploratory capability on the smartwatch.

In this work, we focus on understanding and characterizing components of personal health data queries desired for exploration on the smartwatch. Throughout, our motivation lies in increasing the capability for data exploration directly on the smartwatch. Yet, results conveyed in this work can be seen as being beneficial for a range of devices. To achieve our research goal, we extend and compare to work previously done [31], through the analysis of a publicly available dataset captured in-the-wild [32]. We explore various dimensions, including attributes of data requested, aggregation methods, mechanisms for filtering, and interrogatives used within the queries. By analyzing across these dimensions, we provide a better understanding of *how* people want to explore and access their personal health data on smartwatches. In turn, the results shared can influence interaction in applications catering to smartwatch data exploration.

Our contributions are two-fold: **C1:** an analysis of personal health data queries desired for exploration on smartwatches. Our

MobileHCI '23 Companion, September 26–29, 2023, Athens, Greece

ACM ISBN 978-1-4503-9924-1/23/09.
https://doi.org/10.1145/3565066.3608700

findings provide insight into people's interactive needs when accessing and exploring personal health data on smartwatches. **C2:** a discussion of our findings, centered around the future of personal health data exploration on the smartwatch. We hope our work invigorates discussion and motivates future research, while also providing valuable insights for the interactive design of smartwatch health-related applications.

2 RELATED WORK

2.1 Smartwatch Data Exploration and Interaction

Smartwatch users collect and explore personal health data for a range of reasons. These include, goal and performance monitoring [1, 6], participatory interaction [41], current status viewing, history and trend exploration, and for uncovering actionable insight during an activity [1, 16, 22]. Furthermore, smartwatch users seek unavailable data such as activity-based contexts and preemptive insight [32]. However, it remains difficult on a small-screen smartwatch for health applications to accommodate this broad range of health insight desired [29], given our limited knowledge of the specific querying needs of users. So much so, abandonment due to lack of features and performance is common [12].

As of now, smartwatch health applications and smartwatch visualization research largely focus on glanceable or micro visualizations [1, 4, 7, 17, 27]. These, however, suffer from a lack of interactivity and customization. Users are presented with predefined visualizations and metrics, leaving little room for personalized exploration or the ability to delve into specific and interesting aspects of their data. Importantly, this smartwatch exploration has the potential to take place anytime and anywhere where only quick information needs are required [11]. This limitation in exploration hampers the user's ability to gain meaningful insight within their data. Simply put, this level of exploration on the smartwatch is not enough. Ideally, the analyzed dimensions highlighted within our results can inform broader interaction techniques to expand personal health exploration capability on the smartwatch.

2.2 Natural Language Query Analysis

In this section, we highlight work on natural language query analysis as these often must characterize and dismantle queries for understanding and processing. While we analyze a natural language query dataset in this work, we highlight that we do not propose natural language as the only means of interaction with personal health data on the smartwatch. Instead, we utilize works in this area to simply gain a better understanding of the components involved when querying data and to situate our analysis among related work.

With the increasing capability for devices to perform natural language processing (NLP), even on a smartwatch, many toolkits have become available for use [24, 26]. These toolkits help perform common NLP tasks such as language identification, tokenization, sentiment analysis, named entity recognition, and part-of-speech tagging. However, these toolkits do not explicitly focus on parsing health related information, nor do they offer solution if we do not know what to look for when performing tokenization and part-of-speech tagging. As such, we must first gain a general understanding the components used within potential personal health related queries before using these tools.

Natural language interfaces (NLIs) have become increasingly popular for general interaction (e.g., Siri, Alexa, and Google Assistant have become more pervasive in daily life) [5, 25] and for visual data exploration [21, 37]. While mainly focused towards data experts, research has collected and explored natural language queries across multiple explicit and implicit dimensions. These include data attributes, chart types, data encodings, aggregations, design references, question words, and verb tenses [15, 31, 35, 38, 39]. While not all dimensions are necessary for personal health data querying on the smartwatch (i.e., some of the prior works focus on specific applications such as for visualization creation), these provide insight into required components which we can analyze to handle and process a data-driven query.

3 DATASET AND ANALYSIS

3.1 Dataset

Our work utilizes a publicly available dataset that was collected, through diary study, by Rey et al. [32]. This dataset contains 205 natural language queries desired for exploration of personal health data on the smartwatch. These queries were captured in-the-wild, thus were recorded by participants throughout their daily lives. This has the benefit that the queries reported are less prone to recall-bias and may be increasingly ecologically valid [19]. Within the dataset analyzed in this work, our analysis focuses on the queries themselves while also incorporating other elements of the dataset for granularity (i.e., the relation of the query to a current activity).

For transparency within our work, we highlight details about the collection of the publicly available dataset. Eighteen participants from Canada were involved in query collection. Participants were aged from 18 to 56 ($M = 29.8$) and held a range of occupations. Notably, participants had experience collecting personal health data (on average for 39.3 months) and using a smartwatch (on average for 31.3 months).

3.2 Coding Procedure

In order to analyze the dataset, we coded components of the personal health data queries. To ensure coding accuracy, we followed a procedure used within related works [32, 38]. Specifically, our procedure was as follows: Two researchers first explored the queries independently, creating a coding schema that would outline the potential dimensions that could be assessed. After discussion, dimensions were chosen for which to analyze and code. As a team, a code book was created. Then, the same researchers individually coded a random subset (10% of all queries available) of the data. The assigned codes were then compared for agreement. Until 85% agreement was reached, the code book was refined and a new subset of data was individually coded. Once agreement was obtained, the entire dataset was independently coded using the finalized and agreed upon code book. Finally, any remaining code disagreements were discussed and resolved until a full consensus was reached and a single code was assigned for each dimension explored.

The dimensions explored within the coding schema, and codes used within, followed closely with prior work exploring natural

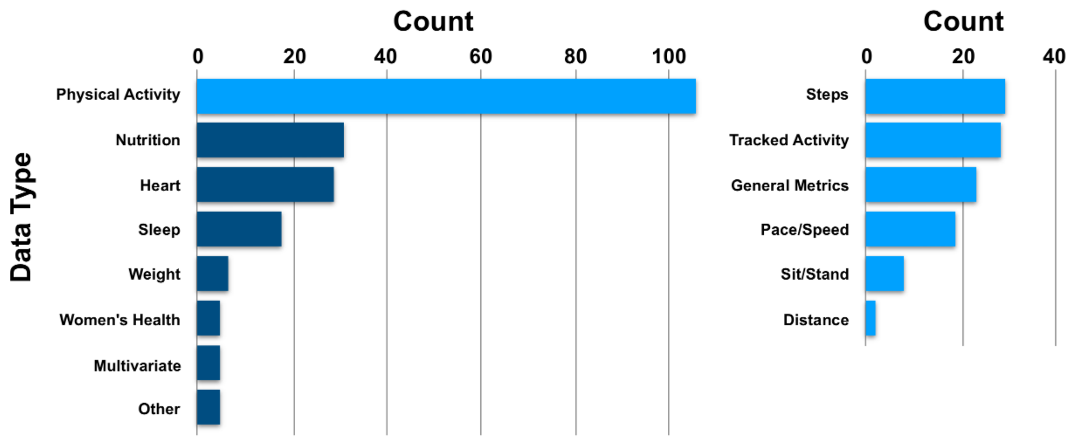

Figure 1: Left: Counts of the data types the queries were focused towards. Right: Sub-codes within the Physical Activity code.

language query interfaces [31, 38]. Hence, we analyze the data type requested, data attributes, filtering mechanisms, and the interrogatives used. Throughout, explicit and implicit/semantic codes are utilized to describe aspects of the data. As the dataset was captured through lay-users, we allow synonyms when using the explicit code rather than deferring these to another code. For example, *peak* can be seen as an explicit aggregation for *max*. In contrast, *how many* is an implicit/semantic aggregation of *count*. These will be highlighted further in their respective subsections below. Lastly, when reporting codes, we provide round brackets containing the count of queries the code captures and the percentage of the dataset this represents.

4 RESULTS

4.1 Attributes of Requested Data

The data requested can be organized into categories, differentiated through activity and the data that is collected. Figure 1 (left) highlights the eight codes used to quantify data types of interest. Not surprisingly, as the smartwatch is generally used as a fitness and physical activity tracker, physical activity data constituted the majority of queries desired (105, 51%). This includes data such as steps, general tracked activities (e.g., "Show me a history of all my dance workouts." or "How many times have I worked out this week?"), and general metrics (e.g., "Did I close all my rings today?" or "How many active minutes am I at?"); see Figure 1 (right) for a complete breakdown. Furthermore, heart data (e.g., heart rate, ECG) (29, 14%) and sleep data (e.g., sleep time, wake time, duration, sleep stages) (18, 9%) were of interest. For granularity, we chose to code heart data separate from physical activity and sleep as this data is captured across and independent of both of the former mentioned codes. Not surprisingly, data that is currently tracked automatically on the smartwatch (i.e., physical activity, heart, and sleep data) were the most queried. In contrast, other data types such as weight, women's health, and nutrition (e.g., number of calories or meals eaten, water intake) which require discrete input or are not currently supported, were less queried. As more data becomes automatically sensed and calculated for tracking, these can be expected to also have relatively higher interest for data exploration.

Attribute references can be seen as words within a query that correspond to a data attribute or specific data point within the collected available data. These references were either (1) Explicit, where the reference in the query was specific to a data point being captured (e.g., "What is my current *heart rate*?" and "How many *steps* did I get during that 2 km walk?") or (2) Implicit, where the reference to data within the query was too broad, could hold different meaning for different people, or required collection of multiple data points (e.g., "Compare my *running stats* from the same time last year" and "Is my *work out better* at my home gym or commercial gym?").

From the 205 total queries in the dataset, we find a large majority, 80% (164) of queries, utilize explicit references to data attributes. Only 20% (41) of the queries showcase implicit referencing. Implicit queries were less immediately data-driven and often contained broad interest into a topic and the need to aggregate data from multiple sources (e.g. "What's the best exercise for me today?"). Queries with implicit referencing of attributes may also present challenges with regard to providing appropriate response. No clear difference in the use of explicit or implicit attribute referencing was seen for queries registered at different times within activity or when away from activity.

4.2 Aggregation

Aggregation references include words that would enable the conducting of a mathematical transform on the data. This is common when performing data analysis (e.g., obtaining the sum, count, average, etc.). Our exploration of aggregations was first coded into the type of aggregation requested; see Figure 2. We found five forms of aggregation and a sixth non-aggregation form: (i) Count (59, 29%) (e.g., "*How many* runs have I completed thus far in 2021?"), (ii) Average (29, 14%) (e.g., "What is my *average* step count per day"), (iii) Min/Max (10, 5%) (e.g., "What was my *fastest* kilometer in my run?"). Other synonyms include: slowest, highest, lowest, peak, best, and worst. (iv) Total (4, 2%) (e.g., "*How many* miles have I *accumulated* through walking, running, and biking over the course of this year"?), (v) Variance (3, 2%) (e.g., "How much has my pace *fluctuated* during my walk"), and (vi) N/A and Current value (97,

		Attributes		Aggregation		Filtering			
		Explicit (e)	Implicit (i)	Explicit (e)	Implicit (i)	N/A	Current (c)	Time (t)	Activity (a)
Yes (107)	Before (24)	20	4	2	22	13	1	6	4
	During (38)	32	6	7	31	9	8	14	7
	After (45)	36	9	13	32	2	6	16	20
No (98)		76	22	15	83	18	19	44	13
Totals		164	41	37	168	42	34	85	49

5 contained both TD and AD

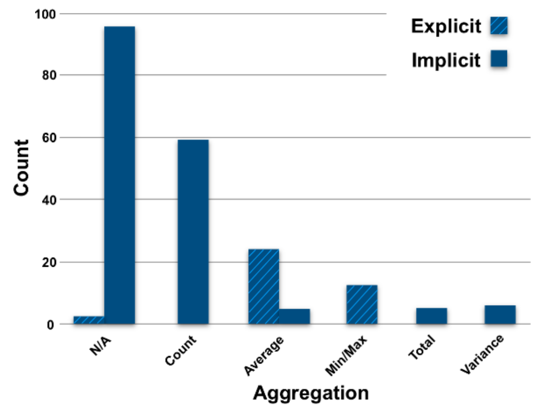

Figure 2: Left: Counts of the codes for attributes, aggregation, and filtering are shown, decomposed by the time within activity the query was collected (as per the dataset by Rey et al. [32]). Right: Specific aggregation references seen within the queries.

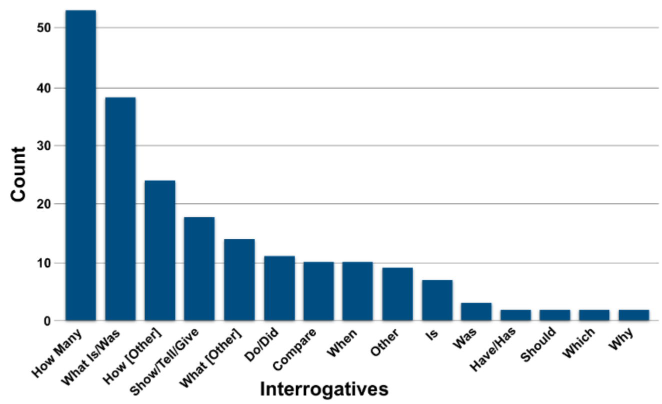

Figure 3: Counts of the interrogatives (elements used to express questions) found within the queries.

48%), where no aggregation is necessary and a value is simply being requested (e.g., "*What is* my resting heart rate?").

We further explored aggregation references through either explicit aggregation, where direct reference to an aggregation transform was used, (37, 18%) (e.g., "What's my *average* walking pace per kilometer" → Average) and implicit aggregation, when phrasing was used, (168, 82%) (e.g, "How long does it take after a walk to get back to resting heart rate?" → Average and "*How many* calories did I burn in the last 4 hours?" → Count). Notably, when we explore the aggregations through this lens, the vast majority of aggregations are performed utilizing implicit requests. When looking at majority, only for average and min/max did people utilize explicit referencing more than implicit referencing.

4.3 Interrogative

Interrogatives, or question words, can provide insight into the aggregation desired. They can also indicate questions versus commands, and hint at appropriate forms of output (e.g., *show me* compared to *tell me*). Figure 3 shows the interrogatives coded. Of interest, over 60% of queries contain either "how" (77, 38%) or "what" (52, 25%) question interrogatives. These can be further broken down into the interrogatives "how many" (53/77), often implying a count aggregation, and "what is/was" (38/52), often implying a single value calculation.

4.4 Filtering Mechanisms

Filtering of data is a common exploratory task (e.g., "Compare walking pace *September and October*" filters the data to the months of September and October while excluding other data). Four codes were used to differentiate filtering mechanisms used within the queries: (i) N/A (42, 20%), where no filtering was needed as the entire data related to the query would be used, (ii) Current (34, 17%) where the current or most recent value would be filtered removing data collected in the past, (iii) Time dependent (85, 41%), where a notion of time was used to filter data (e.g., "What was my highest heart rate in the *last hour*?"), and (iv) Activity dependent (49, 24%), where an instance of an activity is used to filter data rather than an explicit notion of time (e.g., "What was my best kilometer *during my run*?", "Show me my heart rate chart from *today's gym session*.", or "Was the *1st km of my hike* faster than the *last kilometre* today?"). Notably, activity dependent filtering is a subset of time dependent filtering, however is referenced with respect to an activity or activities rather than the specific time frame for an activity.

Time dependent filtering is the most common, as this is a natural means of exploration for personal data. Interestingly, however, the number of queries containing activity dependent filtering shows increase after an activity. This even slightly eclipses time dependent filtering immediately after an activity.

4.5 Components of a Personal Health Data Query

Looking at the results together, four components can be seen to make up a personal health data query on the smartwatch. These include the **interrogative or question word(s)**, the **data subject or attribute(s)**, the **aggregation term**, and the **filtering mechanism**. Importantly, however, not all are needed when querying data. At a minimum, all queries in the dataset contain an interrogative and data subject. This coincides with the results found by Rawassizadeh et al. [31]. Thus, some queries do not contain aggregation or filtering terms, often implying exploration of a current value or of all data captured (e.g., "What is my resting heart rate?"). This is important for interaction of data on the smartwatch, as no matter the interaction modality considered, we must obtain this information at a minimum. Furthermore, we restate that the interrogative can be used as the aggregation term. As such, this is not always explicitly required (e.g., "How many (→ count) steps did I take in the past seven hours?" versus "What was my peak heart rate during my workout?").

5 DISCUSSION

5.1 Lay-Person Exploration of Personal Health Data

We target a broader audience who are interested in exploring and gaining insight from their personal health data. Our analysis is one of the few to focus on lay-person queries. The majority of work in this area has focused on individuals who are familiar with data exploration practices and tools (e.g., data analysts) [20, 37, 38], with to our knowledge only one work analysing lay-person queries for smartwatch data exploration [31].

Within our findings, the influence of lay-person exploration was most evident through implicit aggregation and activity dependent filtering. The majority of queries within the dataset analyzed contained implicit aggregation, a vast difference from the only quarter of queries captured with data experts [38]. While the work done by Rawassizadeh et al. [31] highlights both explicit and implicit notions of time and location, this dimension is not quantified for comparison nor discussed for aggregation. Furthermore, prior work has put focus on time dependent filtering as this is a primary dimension of personal health data [21]; Rawassizadeh et al. also suggest time as the only filtering mechanism [31]. However, activity dependent filtering found in our analysis can be seen as a means for lay-users to more easily recall events performed rather than specific times they were performed. Both these implicit means of exploring data are understandably easier for people, especially on the smartwatch where visual exploration is limited due to the small-screen size and focus may be on a primary in-situ task at hand [3].

5.2 Towards Efficient Input Modalities

Our goal at the outset of this late breaking work, was to better understand components of the personal health data queries desired for exploration on the smartwatch. The question now becomes, how can we create efficient interactions that encompass the knowledge gained from this work?

To support the many facets of data exploration seen in the analyzed queries, natural language (often acquired through speech or keyboard input) offers a solution. Natural language allows for the conveyance of complex queries, has a low barrier in expressing intent, and provides flexibility in phrasing [2, 13]. The above results can be used in future work to create efficient natural language interfaces on the smartwatch. However, while speech is a strong option, it remains to have social acceptability drawbacks [14, 34]. Furthermore, as voice assistants may become more widely accepted in future, voice-based personal data exploration requires for questions to be spoken which can at times hint at underlying values of our personal data and well-being. This is a factor that will be much more difficult to overcome.

Therefore, other input modalities studied for smartwatch use (e.g., bezel interactions, inertial sensing, and efficient touch-based swiping and panning [28, 33, 36, 43, 44]) must be considered and incorporated so that exploration can take place anytime and anywhere. Utilizing knowledge from above, this may include relevant and explicit options and suggestions to tap on (e.g., after tapping the step counter metric seen on screen a "Compare To" option, among others, could appear to further tap on), using the bezel or crown for quick filtering (i.e., from the last walk, to last seven walks, to last 30 walks, etc. or in the case of time dependent filtering from the last hour, to the last day, to the last week, etc.) or filter-mode switching (i.e., from time dependent filtering to activity dependent filtering). These interaction modes will likely need to be combined to fully allow for the handling of query components.

5.3 Large Language Models Versus Logic-Based Approaches

Large language models have become a highly capable and promising tool of late, with their use potentially being able to eclipse the need for some of the results in this work and logic-based interaction approaches. However, deploying large language models on smartwatches for personal health data exploration remains to raise significant concerns which are not easily mitigated. These include ethical concerns surrounding data ownership and control, transparency of model features, and regulatory compliance [18]. Furthermore, large language models can introduce potential bias in responses, lead to over reliance or unsafe use, and exploit user trust to gain private information [42], all of which could be detrimental to personal health outcomes.

As such, this work showcases results which can be immediately used within application and future study for smartwatch personal health data exploration. When we think of our personal data, many of us recognize privacy as something to be conscious of [40, 45]. Through logic-based approaches for interaction, systems can be designed with control, safety, and privacy at the forefront. We strongly encourage readers to consider this as tools become increasingly advanced at enabling broader exploration and analysis of data.

5.4 Limitations and Future Work

This work is limited by the size of the dataset utilized. For reference, the dataset captured by Rey et al. [32] contains 205 queries from 18 participants (on average 11.4 queries per participant). The closest work to this by Rawassizadeh et al. [31], provides a dataset of 716 queries from 131 participants (on average 5.5 queries per participant). With less than 1000 queries combined, it is clear that future work should focus on continually expanding upon the limited available datasets for better analysis and understanding. Yet, study methodology should be carefully considered. In addition to the differences noted in Section 5.1, further diverging results are evident when comparing the ratio of command-based interrogatives (i.e., show, tell, give), where the ratio is greater for participants who had their smartwatch present with them in-the-wild. Considering the differences noted throughout, using an in-the-wild study methodology for data capture, as that done by Rey et al. [32], can provide benefit and should be considered.

In future, we aim to take the results put forward in this late breaking work to realize the potential for interaction with personal health data exploration on the smartwatch. We are excited to utilize this knowledge as we build smartwatch applications to enhance the capability of data exploration for lay users. Interaction will likely take form through multi-modal touch and speech modalities. Multi-modal interaction has been used with success for personal health exploration on smartphones [21]. The results in our work provide knowledge to inform both interactive flow and requirements when enabling these forms of interaction. To this end, we will build and study a natural language parser which can handle entire, and subsets of, queries. Furthermore, we will explore how touch can be used to handle and obtain the required information from a data query. Separately and combined, we can study both through technical evaluation (i.e., query response time and user and system accuracy) and qualitative usability over a period of time.

6 CONCLUSION

This late breaking work provides understanding into queries desired for personal health data exploration on the smartwatch. Through analysis of a public dataset, we identified several dimensions related to these queries, including the requested data types, attributes, aggregation methods, filtering mechanisms, and interrogatives used. The findings emphasize the need for more comprehensive and user-friendly access to a range of health metrics, and the importance of differing data aggregation and filtering options. By considering these insights, practitioners and application designers can develop better applications that meet users' specific needs and preferences for interaction with personal health data. The implications of this research provide valuable guidelines for future works aimed at enhancing the utilization of smartwatches as effective personal health tracking and exploration devices. Considering the findings, we believe that as a research community we can enable individuals to more effectively interact with their health data, thereby promoting better overall well-being and outcomes.

REFERENCES

[1] Fereshteh Amini, Khalad Hasan, Andrea Bunt, and Pourang Irani. 2017. Data Representations for In-Situ Exploration of Health and Fitness Data. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Barcelona, Spain) *(PervasiveHealth '17)*. Association for Computing Machinery, New York, NY, USA, 163–172. https://doi.org/10.1145/3154862.3154879

[2] Jillian Aurisano, Abhinav Kumar, Alberto Gonzales, Khairi Reda, Jason Leigh, Barbara Di Eugenio, and Andrew Johnson. 2015. Show me data": Observational study of a conversational interface in visual data exploration. In *IEEE VIS*, Vol. 15. 1.

[3] Tanja Blascheck, Frank Bentley, Eun Choe, Tom Horak, and Petra Isenberg. 2021. Characterizing Glanceable Visualizations: From Perception to Behavior Change.

[4] Tanja Blascheck, Lonni Besançon, Anastasia Bezerianos, Bongshin Lee, and Petra Isenberg. 2019. Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 630–640. https://doi.org/10.1109/TVCG.2018.2865142

[5] Laura Burbach, Patrick Halbach, Nils Plettenberg, Johannes Nakayama, Martina Ziefle, and André Calero Valdez. 2019. "Hey, Siri", "Ok, Google", "Alexa". Acceptance-Relevant Factors of Virtual Voice-Assistants. In *2019 IEEE International Professional Communication Conference (ProComm)*. 101–111. https://doi.org/10.1109/ProComm.2019.00025

[6] Ana Isabel Canhoto and Sabrina Arp. 2017. Exploring the factors that support adoption and sustained use of health and fitness wearables. *Journal of Marketing Management* 33, 1-2 (2017), 32–60.

[7] Yang Chen. 2017. Visualizing Large Time-series Data on Very Small Screens. In *EuroVis 2017 - Short Papers*, Barbora Kozlikova, Tobias Schreck, and Thomas Wischgoll (Eds.). The Eurographics Association. https://doi.org/10.2312/eurovisshort.20171130

[8] Eun Kyoung Choe, Bongshin Lee, and m.c. schraefel. 2015. Characterizing Visualization Insights from Quantified Selfers' Personal Data Presentations. *IEEE Computer Graphics and Applications* 35, 4 (2015), 28–37. https://doi.org/10.1109/MCG.2015.51

[9] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding Self-Reflection: How People Reflect on Personal Data through Visual Data Exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Barcelona, Spain) *(PervasiveHealth '17)*. Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/3154862.3154881

[10] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1143–1152. https://doi.org/10.1145/2556288.2557372

[11] Jaemin Chun, Anind Dey, Kyungtaek Lee, and SeungJun Kim. 2018. A qualitative study of smartwatch usage and its usability. *Human Factors and Ergonomics in Manufacturing & Service Industries* 28, 4 (2018), 186–199. https://doi.org/10.1002/hfm.20733 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/hfm.20733

[12] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating the Abandonment of Personal Health-Tracking Technologies on Craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 647–658. https://doi.org/10.1145/2750858.2807554

[13] Kenneth Cox, Rebecca E Grinter, Stacie L Hibino, Lalita Jategaonkar Jagadeesan, and David Mantilla. 2001. A multi-modal natural language interface to an information visualization environment. *International Journal of Speech Technology* 4 (2001), 297–314.

[14] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2014. Voice Activated Personal Assistant: Acceptability of Use in the Public Space. In *Human Interface and the Management of Information. Information and Knowledge in Applications and Services*, Sakae Yamamoto (Ed.). Springer International Publishing, Cham, 324–334.

[15] Siwei Fu, Kai Xiong, Xiaodong Ge, Siliang Tang, Wei Chen, and Yingcai Wu. 2020. Quda: natural language queries for visual data analytics. *arXiv preprint arXiv:2005.03257* (2020).

[16] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2018. Activity Tracking in Vivo. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173936

[17] Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, Sean A. Munson, and Marc Hassenzahl. 2016. Exploring the Design Space of Glanceable Feedback for Physical Activity Trackers. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 144–155. https://doi.org/10.1145/2971648.2971754

[18] Philipp Hacker, Andreas Engel, and Marco Mauer. 2023. Regulating ChatGPT and Other Large Generative AI Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1112–1123. https://doi.org/10.1145/3593013.3594067

[19] Gabriella M. Harari, Nicholas D. Lane, Rui Wang, Benjamin S. Crosier, Andrew T. Campbell, and Samuel D. Gosling. 2016. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspectives on Psychological Science* 11, 6 (2016), 838–854. https://doi.org/10.1177/1745691616650285 arXiv:https://doi.org/10.1177/1745691616650285 PMID: 27899727.

[20] Jan-Frederik Kassel and Michael Rohs. 2018. Valletto: A Multimodal Interface for Ubiquitous Visual Analytics. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188445

[21] Young-Ho Kim, Bongshin Lee, Arjun Srinivasan, and Eun Kyoung Choe. 2021. Data@Hand: Fostering Visual Exploration of Personal Data On Smartphones; Leveraging Speech and Touch Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 462, 17 pages. https://doi.org/10.1145/3411764.3445421

[22] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 557–566. https://doi.org/10.1145/1753326.1753409

[23] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-Reflection with Ubicomp Technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, China) *(UbiComp '11)*. Association for Computing Machinery, New York, NY, USA, 405–414. https://doi.org/10.1145/2030112.2030166

[24] Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028* (2002).

[25] Suresh Malodia, Nazrul Islam, Puneet Kaur, and Amandeep Dhir. 2021. Why Do People Use Artificial Intelligence (AI)-Enabled Voice Assistants;. *IEEE Transactions on Engineering Management* (2021), 1–15. https://doi.org/10.1109/TEM.2021.3117884

[26] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations.* 55–60.

[27] Ali Neshati, Fouad Alallah, Bradley Rey, Yumiko Sakamoto, Marcos Serrano, and Pourang Irani. 2021. SF-LG: Space-Filling Line Graphs for Visualizing Interrelated Time-Series Data on Smartwatches. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse; Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 5, 13 pages. https://doi.org/10.1145/3447526.3472040

[28] Ali Neshati, Bradley Rey, Ahmed Shariff Mohommed Faleel, Sandra Bardot, Celine Latulipe, and Pourang Irani. 2021. BezelGlide: Interacting with Graphs on Smartwatches with Minimal Screen Occlusion. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 501, 13 pages. https://doi.org/10.1145/3411764.3445201

[29] Ali Neshati, Yumiko Sakamoto, and Pourang Irani. 2019. Challenges in Displaying Health Data on Small Smartwatch Screens.. In *ITCH*. 325–332. https://doi.org/10.3233/978-1-61499-951-5-325

[30] Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. 2016. Smartwatch in Vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5456–5469. https://doi.org/10.1145/2858036.2858522

[31] Reza Rawassizadeh, Chelsea Dobbins, Manouchehr Nourizadeh, Zahra Ghamchili, and Michael Pazzani. 2017. A natural language query interface for searching personal information on smartwatches. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 679–684. https://doi.org/10.1109/PERCOMW.2017.7917645

[32] Bradley Rey, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. 2023. Investigating In-Situ Personal Health Data Queries on Smartwatches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 179 (jan 2023), 19 pages. https://doi.org/10.1145/3569481

[33] Bradley Rey, Kening Zhu, Simon Tangi Perrault, Sandra Bardot, Ali Neshati, and Pourang Irani. 2022. Understanding and Adapting Bezel-to-Bezel Interactions for Circular Smartwatches in Mobile and Encumbered Scenarios. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 201 (sep 2022), 28 pages. https://doi.org/10.1145/3546736

[34] Christine Rzepka. 2019. Examining the use of voice assistants: A value-focused thinking approach. (2019).

[35] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 365–377. https://doi.org/10.1145/2984511.2984588

[36] Gaganpreet Singh, William Delamare, and Pourang Irani. 2018. D-SWIME: A Design Space for Smartwatch Interaction Techniques Supporting Mobility and Encumbrance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174208

[37] Arjun Srinivasan, Bongshin Lee, Nathalie Henry Riche, Steven M. Drucker, and Ken Hinckley. 2020. InChorus: Designing Consistent Multimodal Interactions for Data Visualization on Tablet Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376782

[38] Arjun Srinivasan, Nikhila Nyapathy, Bongshin Lee, Steven M. Drucker, and John Stasko. 2021. Collecting and Characterizing Natural Language Utterances for Specifying Data Visualizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 464, 10 pages. https://doi.org/10.1145/3411764.3445400

[39] Melanie Tory and Vidya Setlur. 2019. Do What I Mean, Not What I Say! Design Considerations for Supporting Intent and Context in Analytical Conversation. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 93–103. https://doi.org/10.1109/VAST47406.2019.8986918

[40] Lev Velykoivanenko, Kavous Salehzadeh Niksirat, Noé Zufferey, Mathias Humbert, Kévin Huguenin, and Mauro Cherubini. 2022. Are Those Steps Worth Your Privacy? Fitness-Tracker Users' Perceptions of Privacy and Utility. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 181 (dec 2022), 41 pages. https://doi.org/10.1145/3494960

[41] Ryan Vooris, Matthew Blaszka, and Susan Purrington. 2019. Understanding the wearable fitness tracker revolution. *International Journal of the Sociology of Leisure* 2 (2019), 421–437.

[42] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, Zac Kenton, Sasha Brown, Will Hawkins, Tom Stepleton, Courtney Biles, Abeba Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William Isaac, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2021. Ethical and social risks of harm from Language Models. arXiv:2112.04359 [cs.CL]

[43] Hui-Shyong Yeo, Juyoung Lee, Andrea Bianchi, and Aaron Quigley. 2016. WatchMI: Pressure Touch, Twist and Pan Gesture Input on Unmodified Smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 394–399. https://doi.org/10.1145/2935334.2935375

[44] Cheng Zhang, Junrui Yang, Caleb Southern, Thad E. Starner, and Gregory D. Abowd. 2016. WatchOut: Extending Interactions on a Smartwatch with Inertial Sensing. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (Heidelberg, Germany) *(ISWC '16)*. Association for Computing Machinery, New York, NY, USA, 136–143. https://doi.org/10.1145/2971763.2971775

[45] Michael Zimmer, Priya Kumar, Jessica Vitak, Yuting Liao, and Katie Chamberlain Kritikos. 2020. 'There's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information. *Information, Communication & Society* 23, 7 (2020), 1020–1037. https://doi.org/10.1080/1369118X.2018.1543442 arXiv:https://doi.org/10.1080/1369118X.2018.1543442