ABSTRACT

Title of Dissertation:	INTEGRATING HUMAN PERFORMANCE MODELS INTO EARLY DESIGN STAGES TO SUPPORT ACCESSIBILITY
	Benjamin M. Knisely Doctor of Philosophy, 2021
Dissertation directed by:	Monifa Vaughn-Cooke, Ph.D. Assistant Professor Department of Mechanical Engineering

Humans have heterogeneous physical and cognitive capabilities. Engineers must cater to this heterogeneity to minimize opportunities for user error and system failure. Human factors considerations are typically evaluated late in the design process, risking expensive redesign when new human concerns become apparent. Evaluating user capability earlier could mitigate this risk. One critical early-stage design decision is function allocation – assigning system functions to humans and machines. Automating functions can eliminate the need for users to perform risky tasks but increases resource requirements. Engineers require guidance to evaluate and optimize function allocation that acknowledges the trade-offs between user accommodation and system complexity. In this dissertation, a multi-stage design methodology is proposed to facilitate the efficient allocation of system functions to humans and machines in heterogeneous user populations. The first stage of the

methodology introduces a process to model population user groups to guide product customization. User characteristics that drive performance of generalized product interaction tasks are identified and corresponding variables from a national population database are clustered. In stage two, expert elicitation is proposed as a cost-effective means to quantify risk of user error for the user group models. Probabilistic estimates of user group performance are elicited from internal medicine physicians for generalized product interaction tasks. In the final stage, the data (user groups, performance estimations) are integrated into a multi-objective optimization model to allocate functions in a product family when considering user accommodation and system complexity. The methodology was demonstrated on a design case study involving self-management technology use by diabetes patients, a heterogeneous population in a safety-critical domain. The population modeling approach produced quantitatively and qualitatively validated clusters. For the expert elicitation, experts provided internally validated, distinct estimates for each user group-task pair. To validate the utility of the proposed method (acquired data, optimization model), engineering students (n=16) performed the function allocation task manually. Results indicated that participants were unable to allocate functions as efficiently as the model despite indicating user capability and cost were priorities. This research demonstrated that the proposed methodology can provide engineers valuable information regarding user capability and system functionality to drive accessible early-stage design decisions.

INTEGRATING HUMAN PERFORMANCE MODELS INTO EARLY DESIGN STAGES TO SUPPORT ACCESSIBILITY

by

Benjamin M. Knisely

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Assistant Professor Monifa Vaughn-Cooke, Chair
Assistant Professor Mark Fuge
Professor Jeffrey Herrmann
Assistant Professor John Dickerson
Professor Michel Wedel, Dean Representative

Acknowledgements

This dissertation would not have been possible without motivation and support from many different parties. I would like to individually acknowledge some of them here.

First, I would like to thank the dissertation committee members for their participation in this process. They provided valued insight and expertise throughout the process that significantly enhanced the products of this research.

Next, I would like to thank my collaborators at the University of Maryland Medical Center, Dr. Jeffrey Fink and Dr. Lee-Ann Wagner, for their invaluable contributions to this dissertation. Dr. Fink and Dr. Wagner facilitated communication and recruitment of physician participants and provided valuable subject matter expertise. Large portions of this dissertation would have not been possible without their assistance.

I would like to give a special thanks to my academic advisor, Dr. Monifa Vaughn-Cooke. Dr. Vaughn-Cooke was a wonderful mentor. Throughout the process, she provided kind, thoughtful, and open-minded guidance. I believe strong mentorship to be one of the most important elements for success in the pursuit of a Ph.D., and I was lucky enough to have received that.

I would like to thank my peers and colleagues at the University of Maryland. Specifically, I would like to thank graduate students and lab-mates Janell Joyner and Dave Johnson. They both provided invaluable support in this research and made coming into lab a joy. I also want to thank all the undergraduate students at the Hybrid-Systems Integration and Simulation Lab for their assistance over the years.

Last, I would like to thank all my family and friends. Specifically, I would like to thank my parents for their unconditional love and support. They provided the means to make this pursuit a possibility. I would also like to thank my brothers, Andrew and Joey, for their support and for inspiring me to follow this path in the first place. Last, I would like to thank my fiancée Julie Perta for her unconditional love, support, and patience that provided me an environment to succeed in. I am truly blessed to have you as part of my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A1C	Hemoglobin A1C
ANSUR	U.S. Army Anthropometric Survey
AUD	Auditory Discrimination
BIC	Bayesian Information Criterion
BMI	Body Mass Index
CH Index	Calinski-Harabasz Index
CN	Conventional
COPD	Chronic Obstructive Pulmonary Disease
DfHV	Design for Human Variability
DM	Problem-solving and Decision-making
FMM	Fine Motor Movement
GA	Genetic Algorithm
GMM	Gaussian Mixture Model
GUBM	Gross Upper-body Movement
HC	Hierarchical Clustering
HD	Hyper-area Difference
HDL	High-density Lipoprotein
ICC	Intra-class Correlation
ICF	International Classification of Functioning, Disability, and Health
KNW	Applying Existing Knowledge
KL	Kullback-Leibler
MFC	Machine Function Cost
NHANES	National Health and Nutrition Examination Survey
NS	Number of Unique Solutions
PAM	Partitioning Around Medoids
P&C	Physical and Cognitive
PFM	Product Family Member
PR	Proposed
SS	Solution Spread
UD	Universal Design
UFC	Unique Function Cost
VIS	Visual Discrimination

Chapter 1: Introduction

Human beings are highly heterogeneous, varying both physically and cognitively across many dimensions. Human characteristics can have a significant influence on an individual's functional ability to perform tasks associated with a product (Senefeld et al., 2017). Interacting with a product requires users to complete tasks in sequence and simultaneously (Pliner et al., 2021). The functional capabilities of individual users can significantly influence the likelihood that they will successfully perform these tasks.

An individual's physical functioning can differ vastly based on characteristics such as age, level of physical independence, and disease history (Marques et al., 2014; Scheuringer et al., 2005; Senefeld et al., 2017). These characteristics can influence if a user has the capability to perform physical tasks, such as lifting objects or pressing buttons. Similarly, sensory and cognitive functioning are dependent on a wide array of human characteristics (Guilera et al., 2020; Meng et al., 2017; Rudman et al., 2016). Sensation and perception (e.g., visual, auditory, and tactile discrimination) are critical to receive feedback about the state of a system. Cognitive functioning allows humans to perform tasks such as recalling operating procedures and making decisions based on system feedback.

It is critical that users can complete all required system tasks, especially in safety-critical domains, where task failure could result in dire consequences. Examples of user populations who routinely perform safety-critical tasks include patients (Knisely, Levine, et al., 2020), medical professionals (Reddy et al., 2020), transportation workers (Dindar et al., 2020), and the military population (Miranda, 2017).

Leveraging automation in human-machine systems presents the opportunity to alleviate physical and cognitive burden in human users by eliminating or supporting difficult to perform tasks (Bindewald et al., 2014). Automation is increasingly assuming roles once allocated to humans in many product domains, including energy (Oh et al., 2020), healthcare (L. Morelli et al., 2016; Qayyum et al., 2020), and the military (J. K. Proud et al., 2020; Rossiter, 2020). Engineers can customize these automated systems based on the capabilities of the intended user population. To be successful, engineers must consider the trade-offs inherent to expanding system functionality: increasing user accessibility versus decreasing product complexity (cost). This is especially important when designing to accommodate human variability, where a 1-size-fits-all approach may not be appropriate. Variable user capabilities may necessitate additional product offerings, further complicating the accessibility-complexity trade-off.

Despite the need to incorporate heterogeneous user characteristics in design, methodologies for evaluating these trade-offs and customizing product offerings to support human factors objectives are lacking. Further complicating this methodological gap, human factors activities are typically reserved for the late stages of the design process (Irshad et al., 2018, 2019). By delaying these activities, engineers risk discovering new information about the capabilities of end-users that may necessitate expensive and time-consuming product redesign. Performing these activities earlier in the design process could mitigate this risk. This dissertation seeks to provide a methodology to support cost-effective, early design stage product customization to achieve human factors objectives and accommodate human variability.

1.1 Motivation

An inclusive and fair society requires systems that can accommodate the spectrum of user needs present in general *and* specialized populations. Managing human variability presents significant challenges for designers. This dissertation seeks to address some of these challenges, guided by the motivational factors introduced in this section.

1.1.1 Lack of Formative Human Factors Empirical Methods Applied to Heterogeneous Populations

Formative human factors design validation includes efforts to validate the usability of a system in early product design stages. In practice, these efforts are typically performed in-house by individual experts using heuristics or other formal human factors guidelines. To be effective, these efforts should address human variability. Accommodating human variability is not only critical for designing safe and effective systems, it is also required by many regulatory agencies that oversee system development activities (e.g., Department of Defense, 2016; Food and Drug Administration, 2016; National Aeronautics and Space Administration, 2019).

In contrast to formative validation, summative human factors design validation (performed after final design and prototyping) is typically much more rigorous, including recruitment of human users for observational usability studies.

Guidance and best practices are well established for this. There exists little guidance for navigating the challenges associated with performing formative human factors analysis with heterogeneous users. Recruitment of participants for design validation studies can be a significant resource burden from both a cost and time perspective (Christensen et al., 2017; Liao et al., 2015). Further, challenges are often encountered when recruiting from specific, non-general populations. Patient populations, for example, are notoriously difficult to access for lab-based studies (Allsworth, 2015; McHenry et al., 2015). Finally, engineers with limited human factors experience may find practical implementation of these studies challenging. Methodology to navigate these challenges in early design stages is needed.

1.1.2 Lack of Support for "Downstream Neutrality"

Conventional approaches to address human variability primarily exist for use in late-stage design, after physical components and specific interfaces have been designed, contributing to excessive *upstream specificity*. Upstream specificity, a term originated by this research, references design decisions made early in the design process that limit or constrain the design solution space before the problem has become fully understood.

When engineers design a product, they also design (intentionally or otherwise) the functions the product user will fulfill. Conventional approaches ignore the possibility that these underlying functional requirements may be incompatible with user capabilities. Recognizing these incompatibilities in late-stage design, a consequence of upstream specificity, is problematic because products are inflexible and expensive to change. By identifying these incompatibilities earlier in the design process, designers can customize the functional architecture of a product without the cost and burden of redesign. This delayed specification of product elements is coined *downstream neutrality.* The goal of downstream neutrality is to facilitate product customization during the conceptual phase of product design, particularly as it relates to user capability. There is currently a lack of methodology to support practical implementation of downstream neutrality with respect to human factors objectives.

1.1.3 Lack of a Methodology for Segmenting User Populations

Mass customization of products for heterogenous user populations is infeasible in most situations. Grouping users into clusters with similar characteristics can provide targets for engineers wanting to address the needs of densely represented users while remaining viable in terms of the cost associated with additional product variety. In engineering and marketing, this is known as market segmentation and usually involves grouping users with the goal of maximizing group demand for a product (McDonald, 2012). These approaches are typically highly quantitative but treat users as consumers external to the system, not as an integral part whose characteristics can have a notable effect on system performance. In user-centered design, personas are often used to represent archetype or typical user segments (Neate et al., 2019). Persona development is more concerned with human interaction, but most methods are highly subjective and open to significant designer bias. Defining quantitative user segments in early design stages could allow designers to perform targeted differentiation of products with minimal bias. This dissertation proposes a quantitative, user-centered segmentation strategy to support product customization that emphasizes characteristics relevant to product interaction.

1.1.4 Lack of Human Performance Data for Heterogeneous and Specialized Populations

Evaluating and customizing product functionality based on user capability (accommodation) requires human performance data for the intended user population. Human performance data is sparsely available in existing databases and data that is available may not be applicable given the specific design problem. This is especially true in the case of specialized user populations, whose characteristics often deviate from the general population. Further, human performance data is expensive and difficult to collect empirically. Population heterogeneity exacerbates this issue due to the costs associated with recruiting representative population samples. This dissertation explores expert elicitation as a cost-effective and comprehensive means to collect human performance data in heterogeneous and specialized populations to address availability gaps while maintaining feasibility for design organizations of all sizes.

1.1.5 Lack of Models for Product Customization in Early Design Stages

Engineers lack formal models for customizing products to meet heterogeneous user needs in early design stages. In most cases, products can be thought of as a hybrid-system of human and machine elements (hardware and software). In these systems, humans and machines work together to achieve some goal. Function allocation is the process of assigning system functions to humans and machines and could serve as a means for product customization in early design stages.

A modeling approach that facilitates function allocation could support early design stage product customization. Most functional modeling techniques are limited to a product-centric focus. Some recent approaches have adopted a human-machine perspective, however few models exist for use in the early design phases. Those that do exist do not expressly facilitate allocation of functions between humans and machines. In this dissertation, a human-machine system modeling approach is introduced that supports product customization through function allocation.

1.1.6 Lack of Quantitative, User-Centered Tools to Customize Product Families

Allocating system functions between humans and machines presents a trade-off between accessibility and cost. Automating functions decreases workload on the human user by eliminating tasks they would need to perform otherwise. This, however, increases costs associated with design and manufacture of that system. Further, accommodating heterogenous user capabilities requires differentiated products. Product family design is the typical approach to provide cost-efficient product variety. A product family is a set of related products that share elements to attain lifecycle cost benefits while varying other elements to satisfy the needs of particular market segments (Simpson et al., 2001). Product families can serve various user-centered design goals, however little guidance exists to support early-stage product family design activities. The functional architecture of a product could be used to define user-centered product families, however, there exists no quantitative or systematic methodology to do so. This dissertation proposes a multi-objective product family optimization model to configure allocation of system functions to humans and machines, including metrics to evaluate user accommodation and system complexity associated with system functions.

1.2 Overview of Research

The goal of this research is to address the aforementioned research gaps by developing a multi-stage methodology for product family concept generation to support user accessibility. The primary objective of this methodology is to accommodate user capabilities on a population-wide scale while acknowledging the design and manufacturing costs associated with additional product variety. The proposed methodology can be applied to any product that requires human physical and cognitive interaction. The methodology relies heavily on knowledge elicitation efforts using internal medicine physicians: experts who are well versed in the physical and cognitive characteristics of the general population. The method follows this basic pipeline: **Model user groups → Quantify user group performance for system functions → Optimize product family function allocations to maximize accessibility**. The dissertation is split into three mains stages shown in Figure 1 and detailed below along with the stage-specific research contributions:

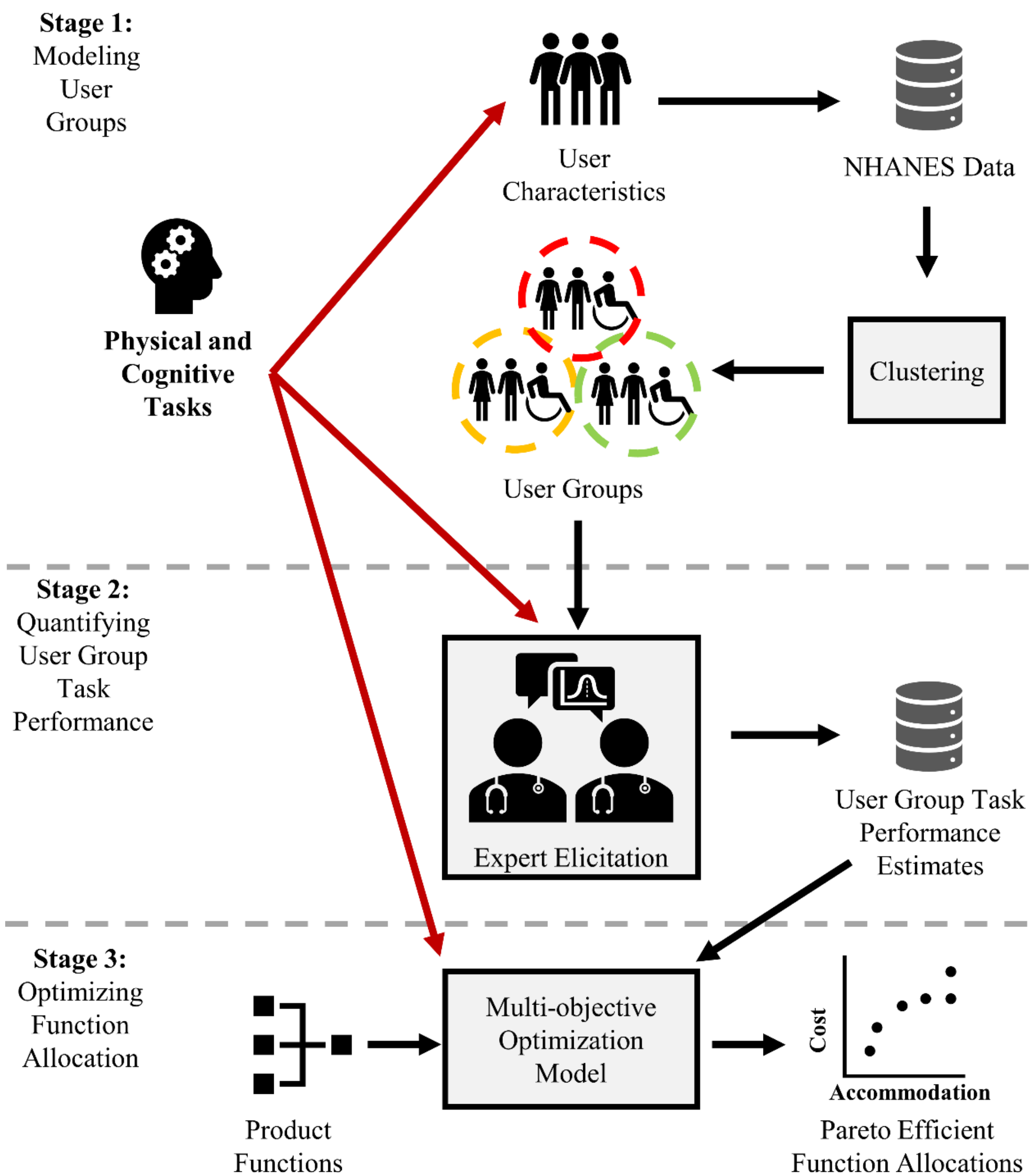


Figure 1: Flow chart demonstrating the relationship between stages of the proposed methodology.

1. **Leveraging Physician Expertise and National Population Data to Model Heterogeneous Population User Groups** – A method for modeling highly prevalent user sub-populations to be used in targeted device personalization is proposed. First, standardized physical and cognitive tasks are generated that

can be applied to any product. They are specifically generated to facilitate physician judgment regarding their performance (coined Physical and Cognitive (P&C) Physician Judgment Tasks), a prerequisite for the subsequent methodological stage. Internal medicine domain-expert input is used to identify user characteristics critical to the performance of standardized physical and cognitive tasks. Variables from the US National Health and Nutrition Examination Survey (NHANES) (Centers for Disease Control and Prevention, 2019) dataset are utilized to quantify the user characteristics. The data are statistically clustered to identify meaningful, task-specific user groups. Task-specific user groups contain users who are expected to perform similarly for a single task. The approach is demonstrated on a diabetes population case study.

<u>Research questions include:</u>

R1. How should designers identify user characteristics critical to product interaction in early design stages?

R2. How can these characteristics be quantified and used to define user groups?

<u>Novel contributions include:</u>

- Generation of standardized physical and cognitive product interaction tasks specifically tailored to facilitate physician judgment regarding their performance.
- A procedure to identify user characteristics critical to the performance of P&C Physician Judgment Tasks.

- A procedure to map user characteristics to NHANES variables for input to statistical clustering.

2. **Quantifying Human Performance for Heterogeneous User Populations using Expert Elicitation -** An expert-driven approach to quantify risk of user error by heterogeneous user populations is proposed. Internal medicine physicians are asked to make quantitative task performance estimates for each task-specific user group identified in the prior stage. Estimates take the form of probability distributions for task success. The approach is demonstrated on a diabetes population case study.

   <u>Research questions include:</u>

   R3. How can quantitative estimates for heterogeneous population task performance be effectively elicited from domain experts?

   <u>Novel contributions include:</u>

   - Detailed procedures and best practices for using expert elicitation to quantify task performance by heterogeneous user populations.

3. **Optimizing Function Allocation for Accommodation of Heterogeneous Populations** – A modeling approach that facilitates function allocation for accessible product families is proposed. Functions from a commonly used function taxonomy (Hirtz et al., 2002) are mapped to P&C Physician Judgment Tasks. Metrics for accommodation and system complexity given allocation of functions are introduced. A multi-objective optimization model is proposed, where function allocations for a family of products are optimally

selected based on maximizing accommodation and minimizing product complexity.

Research questions include:

R4. How can user accommodation and product cost be evaluated in early design stages?

R5. How can product family concepts be optimized for user accommodation and cost?

Novel contributions include:

- An adaptation of a conventional function modeling approach that facilitates allocation of product functions.
- Metrics for evaluating function allocation for user accommodation and product complexity.
- A multi-objective optimization model for allocating functions for a family of products.

In addition to methodological contributions, the methodology is demonstrated on a medical device design case study for the diabetes population, producing data that can be applied for that specific use-case.

Research Questions include:

R6. Can NHANES data be clustered to generate quantitatively and qualitatively separated user groups for the diabetes population?

R7. Can experts produce quantitative task performance estimates for user groups that reflect relative user risk given dominant user group characteristics?

Novel contributions include:

- Task-specific user groups for the diabetes population.
- 27 task performance distributions for the diabetes population.

1.3 Structure of Dissertation

The remainder of this dissertation follows the following format: Chapter 2 contains a literature review of topics relevant to the proposed methods. Chapter 3 introduces a design case study that the method will be demonstrated on throughout the dissertation. Chapters 4-6 will present the methodology, with each chapter corresponding to a stage of the method. At the end of each chapter, the methodology will be applied to the design case study. In Chapter 7, concluding remarks are made for each chapter and the overall contribution of the methodology is discussed.

Chapter 2: Literature Review

This chapter contains a review of literature relevant to this dissertation. Chapter 2.1 discusses approaches to segment heterogeneous user populations into groups for product personalization. Chapter 2.2 presents literature related to quantifying human performance for heterogeneous users. Chapter 2.3 discusses approaches to model systems in early design stages, focusing on function modeling and function allocation. Chapter 2.4 discusses research on product family design and product family design optimization. These topics are linked via the goal of developing a multi-phase design methodology for generation of accessible product family concepts in early design stages.

2.1 Segmenting User Populations

In this section, several topics related to segmenting users into user groups are discussed. First, contrasting perspectives from several disciplines on segmenting users are discussed. Then, the utility of those approaches is discussed in the context of personalized design.

2.1.1 Perspectives on User Segmentation

Stratification of individuals into groups is a common practice when designers want to customize a product or a system to address varying user wants or needs (Pallant et al., 2020; Wedel & Kamakura, 2012). Groups provide targets for product variants while not necessitating every product be individualized. For a given population, the maximum number of segments that can be defined is equal to the number of distinct users in the population (assuming membership is exclusive). In most cases, it is impossible create a completely customized product for this theoretical maximum number of groups. Instead, segments are defined based on characteristics that represent large numbers of users such to maximize generalizability of the segments but minimize the number of segments that must be represented (Tipton & Matlen, 2019).

In typical practice, populations are segmented based on demographic or geographic variables. This is primarily due to the ease of collecting this data. This may not be sufficient when designing for maximum human performance (Privitera, 2020). In general, these variables are not directly causally linked to human performance. For example, an elderly individual may have difficulty performing a mobility task, however, it is not because they are elderly that they struggle. It is because as an elderly person, they are more likely to have some disease or under-lying condition that influences mobility. Grouping users based on characteristics with closer links to capability provides more meaningful segments.

In user-centered design, personas are often used to represent the archetype or typical user (Neate et al., 2019). Oftentimes, several personas are defined for a user population, and represent the central tendencies of prominent user segments. Personas are researcher-based, descriptive models of the typical user(s) of a given system. They are used to ensure that assumptions about intended users are explicitly stated and understood across a design team (Korsgaard et al., 2020). Personas have been used to segment based on user demographics, goals, attitudes, motivations, identity, relationships, affiliations, and design preferences (Carey et al., 2019; Warin et al., 2018). Persona development is typically a highly subjective process, and often relies on qualitative methodology, and therefore may be open to substantial bias. Common techniques for generating personas include interviews, fields studies, usability tests, and ethnography (Salminen, Santos, et al., 2020), though quantitative and mixed methods are becoming more common. Clustering algorithms (e.g., k-means, hierarchical, etc.) are popular for identifying personas from unstructured user data (Mesgari et al., 2015; Tanenbaum et al., 2018; L. Wang et al., 2018). Other techniques popular in quantitative persona development include principal component analysis (L. Wang et al., 2018), latent semantic analysis (Miaskiewicz et al., 2008), and non-negative matrix factorization (An et al., 2017). Additionally, emphasis has been placed on extracting data from web platforms and other sources of "personal big data" (Salminen, Guan, et al., 2020).

In some marketing and engineering contexts, segmentation of users is often referred to as market segmentation. Market segmentation and persona development are distinct but overlapping strategies. Segmentation in this context generally differentiates itself by focusing on attributes that predict purchasing behavior and maximizing demand, as opposed to user performance (Sherkat et al., 2016). Purchasing behavior and user performance are related but not perfectly correlated outcomes (i.e., the design that maximizes product demand may not maximize user performance). Quantitative methods are more common in this context, often taking advantage of various approaches to statistically clustering users (Bose et al., 2020; Ma & Kim, 2016; Ramasubbareddy et al., 2020). Examples of variables used to quantitatively segment populations include demographics, preferences, personality, and purchase history (Paço & Raposo, 2010; Pallant et al., 2020; Pomarici et al., 2017). Segmentation is also achieved using attributes of existing products, for example with design attributes for an aerodynamic particle separator (Tucker et al., 2010) and for a universal electric motor design (Ma & Kim, 2016).

Some quantitative approaches to identify user groups based on user-product interaction have been proposed. For example, math-based performance simulations based on user and contextual variables were developed to evaluate design solutions for varying user segments (Bekhradi et al., 2015; Yannou et al., 2013). Brolin et al. (2016) utilized summary statistics of characteristics related to user capabilities to generate and cluster a synthetic population of users to be used as design targets. Product usage data has also been used to group individuals based on patterns of behavior, examples including clustering of smartphone users (Razavi, 2020; Zhao et al., 2020), and classifying user activity based on shoe-embedded sensors (Ghosh et al., 2016). Links to task-specific characteristics of the user are limited in these methods, making personalization difficult without knowing what led to the variability of interaction.

2.1.2 Personalized Design

The aforementioned methods for user segmentation have been used widely in design personalization. Personas are commonly applied in the design of software (Andriella et al., 2018; Anvari et al., 2017), hardware (Alsager Alzayed et al., 2020; Stevenson et al., 2018) and service systems (Idoughi et al., 2012). The benefits of personas in the context of design are numerous. Miaskiewicz and Kozar (2011) identified and had experts rank 21 different ways personas are used in the design process. The top benefits identified were focusing product development on users instead of specific technological limitations or opportunities, guiding prioritization of product requirements, and prioritization of the needs of the most important stakeholders or users. Viana and Robert (2016) described the benefits of using personas for user-interface design, including facilitating communication about the goals of the product within a design team, challenging organizational assumptions about the user, creating empathy towards the user, and as a surrogate in user testing.

Market segmentation strategies have also been integrated within product design in what is known as market-driven design, where the expected demand of a product is maximized amongst other engineering constraints (Donndelinger & Ferguson, 2017). In this context, product families have been proposed as a cost-effective means for customizing products to the needs of consumer segments. Product families provide additional product variety while taking advantage of economies-of-scale by standardizing certain product elements across product family members (Otto et al., 2016). Core product features are shared across user groups, while other features are varied or added to meet specific wants or needs. Researchers have sought to provide methods for optimally designing product families to efficiently meet user needs (Ma & Kim, 2016; Simpson et al., 2001; Sinha & Suh, 2018; Q. Wang et al., 2019). While highly quantitative, these methods primary focus on maximizing demand, which does not necessarily align with maximizing usability or human performance.

Market segmentation also appears when implementing software and multimedia. Breaking users into groups and tailoring software to the needs of the group can improve user satisfaction, and encourage software acceptance (Sherkat et al., 2016). Segmentation is important when pricing software to maximize demand (August et al., 2019). Segmentation strategies have also been implemented into existing software and website platforms to support effective recommendation systems for users (Bose et al., 2020).

Recent work on adaptive design suggests that it may not always be necessary to customize individual products to user segments. Adaptive interfaces monitor the state of the system and the state of the user to adapt the display and available user actions to maximize interaction performance in real-time (Lavie & Meyer, 2010). In some cases, these adaptive systems may negate the need for entirely separate products if features critical to interaction can be automatically adapted to the needs of the specific user. Advanced systems have begun to take advantage of physiological data from biosensors, such as eye-trackers, heart-rate sensors, and motion trackers, to adapt systems to the needs of the individual (Çığ Karaman & Sezgin, 2018; Georgiou & Demiris, 2017; Lin et al., 2017). Other systems have taken advantage of the vast amounts of data available from social media platforms to provide customized services (Jeon et al., 2020). Adaptive automation has also been utilized to dynamically allocate functions between humans and machines (Bindewald et al., 2014). Other examples of adaptive design include adaptive cyber-security based on user characteristics and behavior (Addae et al., 2019), adaptive interface multimodality for interactive devices (Kong et al., 2011), design of an intelligent adaptive interface for control of UAVs (Ilbeygi et al., 2019), and user-adapted e-learning platforms for students (Santoso, 2021). While most adaptive design focuses on software platforms and interface design, some adaptive design features have been explored in the hardware context (Jevtić et al., 2019; Zhou & Liu, 2020).

2.2 Quantifying Heterogeneous User Performance

Quantifying human performance for user groups can reveal distinct group requirements to drive product customization. This section presents several perspectives to quantify human performance for use in design. First, simulated-use testing is presented as the go-to human factors approach to quantify human performance heterogeneity. Discussed next is design for human variability, a model-based engineering perspective. Last, expert elicitation is presented as a cost-effective alternative to quantify human performance.

2.2.1 Simulated-use Testing

Human factors evaluation and simulated-use testing are used to provide evidence that a design conforms to the needs of the intended user population (Barnum, 2020). Simulated use testing seeks to replicate the conditions of system use, providing researchers the opportunity to observe user behavior in a controlled lab setting. There are many benefits to this practice. Simulated-use testing allows researchers to collect quantitative data on user performance and user error while controlling confounding factors (Liao et al., 2015). This user data can be incorporated into the design process and support decision-making to mitigate downstream user error, thus supporting user safety and preventing potential recalls and litigations (Johansen, 2018).

A variety of metrics can be used to quantify performance during simulated use-testing. Objective measures typically include accuracy, task timing, or error rates (Claypoole et al., 2019; F. Morelli et al., 2017; Radwin et al., 2014). Neurophysiological data can also be used to supplement these measures, examples including heartrate variability (Delliaux et al., 2019) and pupil response (Van Acker et al., 2020). While useful, the cost to acquire this data can be high and may require specialized equipment. Subjective measures of performance can be used to supplement direct observation. Self-reported data is generally less burdensome to elicit from participants, however it may be subject to various biases that objective metrics are not (Rosenman et al., 2011).

Regulatory agencies have indicated that human factors evaluation is critical to design safe and effective systems across industries. In Table 1, key quotes regarding human factors testing and heterogeneous populations have been isolated from regulatory standards and guidance. These standards and guidance demonstrate the importance of representing heterogeneous users in design validation. The terms "intended users," "representative users" and "user populations" are referenced repeatedly among other related terms. Despite the stated need to integrate user variability, there is little guidance on how to successfully incorporate these requirements into the design process.

Table 1: Excerpts from publicly available standards and guidance on representing hetergeneous users in design validation.

Organization	Document	Key Quotes
US Food and Drug Administration (FDA)	Applying Human Factors and Usability Engineering to Medical Devices (FDA-2011-D-0469) (Food and Drug Administration, 2016)	**Section 8:** "Human factors validation testing is conducted to demonstrate that the device can be used by the intended users without serious use errors or problems, or the intended uses and under the expected use conditions."
		Section 8.1.1: "The human factors validation test participants should be representative of the range of characteristics within their user group…" "…if different user groups will perform different tasks or will have different knowledge, experience, or expertise that could affect their interactions with elements of the user interface and therefore have different potential for use error, then these users should be separated into distinct user populations."
US National Aeronautics and Space Administration (NASA)	NASA Spaceflight Human-System Standard (NASA-STD-3001) (National Aeronautics and Space Administration, 2019)	**Section 4: "**A systems engineering process that adequately considers human performance variability and limitations during spacecraft design, development, testing, and evaluation is of critical importance to the health, safety, and performance of flight crews, as well as to the protection of hardware and systems."
		Section 10.1.1.1: "Usability testing must be part of the verification process."
	Human Integration Design Handbook (NASA/SP-2010-3407) (National Aeronautics and Space Administration, 2014)	**Section 3.3.8.5:** "It is imperative to use representative users in the simulations and evaluations to ensure that results capture the capabilities of the user and are relatable to the mission situations."
US Nuclear Regulatory Commission (NRC)	Human Factors Engineering Program Review Model (NUREG-0711) (OHara et al., 2012, p. 0711)	**Section 11.4.3.4:** For validation testing, "To properly account for human variability, the applicant should use a sample of participants that reflects the characteristics of the population from which it is drawn. Those characteristics expected to contribute to variations in system performance should be specifically identified…"

Table 1 (continued)

US Department of Defense (DoD)	Human Engineering Requirements for Military Systems, Equipment, and Facilities (MIL-STD-46855A) (Department of Defense, 2016)	**Section 5.3.2:** Planned test and evaluation shall include "Use of personnel who are representative of the range of the intended user populations in terms of aptitudes, skills, capabilities, experience, size, and strength; wearing suitable clothing and equipment appropriate to the tasks (use of personnel from the intended user population is preferred)"
	Defense Acquisition Guidebook (DAU, 2020)	**Chapter 5-4.2.2.2:** "The PM [Program Manager] should use a truly representative sample of the target population during Test and Evaluation (T&E) to get an accurate measure of system performance. A representative sample during T&E helps identify aptitude constraints that affect system use."
Federal Aviation Administration (FAA)	Human Factors Design Standard (DOT/FAA/HF-STD-001B) (Ahlstrom, 2016)	**Chapter 6.1:** "The result of using this document in development and acquisitions will be a more usable system. However, even systems that are carefully designed using this document in conjunction with a human factors expert will need to be verified through means such as prototyping and testing with representative users."
National Institute of Standards and Technology (NIST)	Human Engineering Design Criteria Standards Part 3: Interim Steps (for the Department of Homeland Security (DHS)) (Furman et al., 2014)	**Executive Summary:** "…systematically adopting and applying HIS [Human System Integration] criteria within DHS will be a challenge because of the department's large and extremely varied user population." "DHS needs to conduct some type of user acceptance and usability testing of potential new technologies before deploying them in the field."

2.2.2 Design for Human Variability

Design for Human Variability (DfHV) is a quantitative, model-based design for X approach to evaluate human performance. DfHV focuses on quantifiable, physical characteristics of the user, most often related to anthropometry (Ferguson et al., 2015). Anthropometry is a measurement science that focuses on the human body (Heymsfield et al., 2018). DfHV seeks to optimize products and systems for safety, fit, and performance for broad populations using statistical modeling and ergonomic/human factors tools (Garneau et al., 2014). Accommodation, a measure of the portion of target users able to use an artifact in the desired manner, is usually the objective to be optimized (Boyd & Parkinson, 2015).

Typically, DfHV starts by identifying the user population, determining the variables that may influence user-artifact interaction, and quantifying the relevant human variability. Variability is quantified using existing data or with model estimations (Garneau et al., 2014). Databases, such as the US Army Anthropometric Survey (Gordon et al., 1989), contain anthropometric measures collected from a population of interest and can serve to represent variability of human form.

Surveying the desired user population is also an option, however, it can be expensive and is limited by participants available for recruitment. A third option, which is an active area of research, is to estimate population characteristics using statistical models. Parkinson & Reed (2010) introduced a technique to synthesize virtual user populations from databases for assessment of accommodation, utilizing principle component analysis and linear regression**.** Similarly, Brolin, Högberg, Hanson, & Örtengren (2017) introduced an adaptive regression-based methodology to synthesize population data. Examples where synthesized populations have been used to inform artifact design include multi-user workstations (Mahoney et al., 2015), prosthetic heart valves (Aycock et al., 2015), and cockpit seat design (Poirson & Parkinson, 2014).

Designing for variability is typically accomplished by specifying adjustability or sizing artifacts appropriately (Garneau et al., 2014). A virtual fitting trial is used to assess the fit between an individual and an artifact using synthesized population measures or using digital human models that perform the fitting trial in a virtual graphical environment (Garneau & Parkinson, 2011; Godwin et al., 2007; Mahoney et al., 2015). Boundary manikins, individuals representing the extremes of the population, are usually sufficient for assessing accommodation (Boyd & Parkinson, 2015).

While highly objective and relatively easy to apply, DfHV has several limitations. DfHV methods utilize synthesized populations for assessing product fit. These synthetic populations often rely on databases that are out of date, do not sufficiently represent the population at large, or may not contain the desired measures (Nadadur et al., 2016). For example, the oft relied on 1988 US Army Anthropometric Survey (Gordon et al., 1989) does not adequately represent the shifting demographics and form of the US population at large, or even the current US army population (Garneau et al., 2014). Described by de Vries & Parkinson (2014), disproportionate disaccommodation refers to a design that does not proportionally accommodate the needs of certain sub-groups, and is both an ethical and performance concern. This is especially a concern to demographic minorities, whose needs may unintentionally go overlooked in conventional design settings. In addition, DfHV methods typically only consider physical variability. These methods neglect to address the variations in user cognitive and functional capabilities that can also have an impact on user-product interaction.

2.2.3 Expert Elicitation

One way to overcome challenges associated with recruiting heterogeneous users is to supplement recruitment-based studies with other approaches. Expert elicitation is the process of eliciting judgments regarding the value of unknown quantities, often in the form of a probability distribution, from individuals who have been judged to be experts (Brownstein et al., 2019). Expert elicitation has previously been proposed as a means to quantify human performance (P. Liu et al., 2020; Pandya et al., 2020). In one of the most popular implementations, the Cooke protocol, several experts are asked to individually estimate quantities of interest as well as their uncertainty of the estimate (typically as percentiles). Estimates are performance weighted and pooled to achieve a robust consensus (Colson & Cooke, 2018; Cooke, 1991). Other protocols require experts to collaborate on estimations, such as with the IDEA protocol. The advantage of having experts convene is that it can help clarify linguistic ambiguity and promote critical thinking. The disadvantage is that requiring experts to convene can be logistically difficult and expensive (Hemming et al., 2018). There are other available protocols (e.g., Delphi method (Skulmoski et al., 2007), SHELF (Gosling, 2018)) in literature and selecting the correct one will depend on the resources available and the availability of domain experts.

The advantage of expert elicitation over other approaches discussed is the low resource burden and the flexibility (Hanea & Nane, 2019). When quantifying heterogeneous human performance, expert elicitation minimizes the need to recruit users. Further, there are no constraints on what quantities can be elicited. This said, the limitations of expert elicitation are significant (Morgan, 2014). Ultimately, without additional validation efforts, there is no way to guarantee estimations are sound even when procedures are followed meticulously. This is a highly subjective process, subject to various cognitive biases (“anchoring”, “range-frequency”, and overconfidence for example), but significant prior research has been performed to provide guidance to make the process as objective and scientific as possible (O'Hagan, 2019). Ultimately, expert elicitation provides a subjective but cost-effective means to elicit unknown quantities when other options are infeasible (Hemming et al., 2018).

2.3 Modeling Product Function

Customizing a product in early design stages requires an abstraction, or a model of the product. This chapter discusses function modeling, a process that seeks to represent a system as solution-neutral elements that describe what the system does without describing physical product elements. Attempts to incorporate human factors into models of function are discussed as well.

2.3.1 Function Modeling

Function modeling is a analytical early-design stage process to explore the design solution space for potential concepts (Patel et al., 2020). Function models seek to provide a solution-neutral (i.e. with no physical embodiment) representation of an artifact or system that represent what that system does, it's purpose and it's behavior (Tomiyama et al., 2013). It is a way to formalize the understanding of a system, to support the generation of new product concepts, and to support the analysis and improvement of existing products and systems (Mokhtarian et al., 2017). Function modeling is useful because it helps to represent the purpose of the artifact, explain behavior or structure, capture functional customer requirements, and illustrate an overview of the artifact (Tomiyama et al., 2013)

Some of the most common approaches to modeling function are based on the flow-based thinking sometimes referred to as function structure (Pahl et al., 2007). These approaches represent material, energy, and information flows and the transformations they undergo through a system (Yildirim et al., 2017). Function structures are typically represented as a flow chart with functions represented as blocks and flows represented as arrows connecting functions. Other well-known approaches include the function-behavior-structure model (Umeda et al., 1996) and the structure-behavior-function model (Goel et al., 2009), which link the function, the behavior, and the structure of the system in a single representation.

Language plays an important role in appropriately conveying function, and in creating effective models. Function often appears in engineering as natural language or in subject-verb-noun triplets (Tomiyama et al., 2013). The functional basis (Hirtz et al., 2002), a common language for describing product and system functions, is one of the most popular function taxonomies and is used to reduce ambiguity and facilitate communication between modelers (Stone & Wood, 2000).

2.3.2 User-Centered Perspectives

There have been several attempts to incorporate human factors considerations into models of function (Sun et al., 2018). Ramachandran, Caldwell, & Mocko (2011) proposed the function interaction model, which includes an abstracted representation of the user and user activities. It was demonstrated to be a more effective tool for concept generation than standard function modeling. Sangelkar et al. (2012) introduced the action function diagram, a variation on the function structure diagram that links user activities to product functions. It highlights the differences between universal products and typical products and can be used to reveal heuristic rules for universal design. This model took advantage of the World Health Organization International Classification of Functioning, Disability, and Health (WHO ICF) (World Health Organization, 2001) formal classification language for describing interaction. Similarly, Soria Zurita et al. (2020) linked modes of human failure with functions.

Affordance-based methods have also been used to link the user to product functions. In the context of engineering design, an affordance is a relational benefit that an artifact offers an individual and is an emergent property of the user-artifact system (Cormier et al., 2014; Galvao & Sato, 2005). While not completely agreed upon, Maier & Fadel (2009) stated that affordances have the following properties: 1) Complementary – affordances cannot exist in either subsystem or in isolation; 2) Polarity – affordances can be positive or negative; 3) Multiplicity – a system can have multiple affordances; 4) Quality – affordances can have varying quality; 5) Form dependence – affordances depend on the artifacts physical structure.

Affordances have been shown to be a useful construct for capturing and modeling user needs in early design and redesign phases, aiding in producing more usable and desirable products. Galvao & Sato (2005) were one of the first to demonstrate this idea by linking product functions to user tasks in a matrix form and using this to generate design solutions based on corresponding functional affordances. Maier & Fadel (2009) introduced the Affordance Structure Matrix, another architecture-matrix representation, which links affordances to physical product components. Further efforts to incorporate affordances into product development include the development of an affordance basis (Cormier, Olewnik, & Lewis, 2014), reconciling function and affordance representations (Ciavola et al., 2015), and connecting affordance to design for environmentally conscious behavior (Srivastava & Schumann, 2013).

While the previous efforts are useful for identifying where human interaction is necessary during product use and the nature of the interaction involved, they do little to identify the specific elements of the human involved in the interaction. Similar to functional representations of artifacts, a human interaction can also be abstracted to a functional classification. Cage (2017) proposed a standardized approach for mapping musculoskeletal interfaces to product components and functionally classifying the interaction from one of several generic interactions. This approach is used to identify physical human parameters important for accommodating musculoskeletal variability as well as identifying the human functions critical for successful product use.

2.3.3 Function Allocation

A product and a human product user form a human-machine system, where the product and the human perform various functions to achieve an overall goal. Function allocation, the process of distributing functions or tasks within these systems, is often discussed in human factors research but typically less represented in engineering design methodology (de Winter & Dodou, 2014). It seeks to answer the question "Who does what in this system?" and typically takes place in the conceptual design phase, and serves as a basis for the machine logic of the system (Feigh & Pritchett, 2014). Developed in the 1950s, the Fitts Lists was one of the first attempts to formalize allocation of functions to humans and machines. The Fitts list, as well as the many lists it inspired, was a static list of strengths and weaknesses of both human and automation to be used as a basis for allocating functions (de Winter & Hancock, 2015). Similar lists contained "levels of automation", which specify degrees to which control of a task is given to a human or machine (Endsley & Kaber, 1999; R. W. Proud et al., 2003). Though it was foundational work, the Fitts list and other "MABA-MABA" (Men Are Better At, Machines Are Better At) works have been met with significant criticism for being static, impractical, and for having 1-dimensional criteria for automation (Fuld, 2000; Hancock & Scallen, 1996; Sheridan, 2000). Critics have also pointed out that 1-to-1 substitution of human with machine functions is flawed because automation often results in emergent properties due to human-machine interaction (Dekker & Woods, 2002).

Due to automations increasing presence in most domains, function allocation has received renewed interest in recent years and many researchers have attempted to reconcile prior criticisms. Feigh & Pritchett (2014) outlined key criteria for effective function allocation, including: 1) Each agent must be allocated functions it is capable of performing; 2) Each agent must be capable of performing its collective functions; 3) The function allocation must be met with reasonable teamwork; 4) The function allocation must support the dynamics of work; and 5) The function allocation should be a result of deliberate design decisions. These requirements then went on to inform accompanying modeling (Pritchett et al., 2014b) and measurement (Pritchett et al., 2014a) frameworks for allocation of functions.

Adaptive and contextual allocation of functions has also become a subject of interest. Adaptive automation refers to systems in which the allocation of certain functions change with time (Sheridan, 2011). Though often considered impractical, advancements in sensing and methods for exploiting sensor data has made adaptive automation increasingly feasible (Feigh et al., 2012; Mannaru et al., 2016). Attempts have been made to model and apply dynamic function allocation. Bindewald, Miller, & Peterson (2014) demonstrated a modeling framework that supports the dynamic allocation of tasks in a computational work setting. Kidwell, Calhoun, Ruff, Parasuraman, & Mason (2012) successfully applied adaptive automation to the control of multiple autonomous vehicles simulation.

Several authors have suggested the need to address mismatches in responsibility, ability, and authority in function allocation (Kaber, 2018; Pritchett et al., 2014b). A machine agent may be given the authority to execute a certain function, but a human agent often still has the implicit responsibility over the outcome of said function. Another example of this type of mismatch is when humans are given authority and responsibility over functions that they do not have the ability to perform. Put simply, responsibility should not exceed ability, and should not exceed authority (Flemisch et al., 2012). Responsibility that exceeds either of these can result in deficiencies in human performance, prevent effective cooperation, and potentially erode at the trust between human and machine agents.

In heterogenous populations, customizing the allocation of functions for distinct user groups could mitigate the risk of detrimental mismatches between user capability and user responsibility. Product family design could provide a cost-effective means to achieve this.

2.4 Product Family Design

In this section, an overview of product family design methodology is discussed. Then, attempts to formalize product family design as an optimization problem are presented.

2.4.1 Product Family Design Overview

A product family is defined as a group of related products that share features, components, or subsystems to attain lifecycle cost benefits while varying other elements to satisfy particular market niches (Simpson et al., 2001). The product family refers to the set of products that share elements. The product platform refers to the elements that are shared between products from which product variants can be derived (Gauss et al., 2021). Variety is the diversity of products within a product line (Jiao et al., 2007). Typically, variety is achieved in two ways: 1) scaling of elements; or 2) swapping/adding functional elements. Scale-based platforms derive variety by changing the element parameters with the same functional capacity. Functional-based platforms derive variety by configuring function-based modules. The process of creating this variety to match the target population is product positioning (Jiao et al., 2007).

While there are many benefits to product family design, the two core advantages are related to engineering effort and manufacturing complexity (Gauss et al., 2021). By encouraging a modular product design, engineers can reuse technical solutions, reducing design resources and eliminating redundancy in design solutions. Further, product family strategy can reduce the number of manufacturing lines required, eliminating upfront costs (Fiorineschi et al., 2014).

Product family design often focuses on physical product architecture in the mid- to late-stages of design, however, there some examples of product family strategy being implemented at the functional stage. Function structure heuristics refers to a set of clustering rules to be applied to a functional representation of a product or system. Typically, the system of interest is represented as a function structure, a diagrammatic representation of material, energy, and signal flows and the functional transformations they undergo to achieve a goal (Fiorineschi et al., 2014).

2.4.2 Product Family Optimization

After a product family architecture has been defined, inclusion/exclusion and scaling parameters of product elements can be optimized. Product family optimization has been a popular area of research for some time. There are many variables and objectives to consider in these problems, however Pirmoradi, Wang, & Simpson (2014) identified 3 common classes of problem: 1) The platform configuration is known and the optimum design variables for that configuration is the objective (Michalek et al., 2005; Wäppling et al., 2011); 2) The optimum configuration of products (number of family members, configuration of modules, etc.) is the objective (Akai et al., 2010; Fujita et al., 2013); and 3) Simultaneous consideration of both (Ma & Kim, 2016; Pirmoradi et al., 2015).

Multi-objective optimization is a class of optimization problem where multiple competing objectives are considered simultaneously. It is a popular approach to product family optimization. Multi-objective optimization approaches can integrate engineering design, customer values, production cost, and other product family objectives into a single problem (Unal et al., 2017). These methods identify ranges of product family designs that are determined to be non-dominant, or pareto-efficient, based on the chosen criteria (Simpson et al., 2012). Akai et al. (2010) considered deployment of product family modules, optimizing based on maximizing profits and minimizing consumption of engineering resources. Sinha & Suh (2018) developed a model for minimizing structural complexity of a given design and maximizing the degree of modularity. Non-financial and engineering objectives are often considered as well, for example, when examining trade-offs between environmental impacts and product costs (Kim & Moon, 2017; Q. Wang et al., 2019).

Chapter 3: Case Study Description – Diabetes Self-Management Technology

The methodology proposed in this dissertation will be demonstrated on a design case study. The design case study was selected based on its appropriateness for the method, which is dependent on the characteristics of the target user population. The two criteria for judging appropriateness are listed below. The first criterion is a hard requirement. The second criterion is not required, but the potential benefits provide additional justification for using the method. The criteria are:

1. **Heterogeneous Users –** The population is heterogeneous with respect to characteristics that influence product interaction. The assumption is that heterogeneity of these characteristics will result in heterogeneous functional design requirements. Heterogeneity is relative and will need to be judged based on prior knowledge about the characteristics of the population as well as the frequency and variability of usage-related issues experienced by the population (therefore requiring varied design requirements).

   Heterogeneous populations can be associated with consumer goods, where the user population is the general population, and includes products such as cell phones and laptops. Heterogeneous populations also appear for specialized products. This includes untrained populations, such as medical device users, whose characteristics can be widely varied. It can also include trained populations, such as occupation-based populations who require specialized tools. These populations will typically be relatively less heterogeneous as there are often barriers-to-entry that will limit the variance of the population.

2. **Safety-critical or highly regulated domain –** The health and well-being of a system stakeholder is dependent on the ability of the population to appropriately interact with the product, or the product domain is highly regulated with respect to meeting the usability needs of the intended user population. In most cases, this requirement and the prior requirement will coincide, as safety-critical domains (e.g., transportation, energy, defense, healthcare) require comprehensive regulation. Since this method will add time/resources to the design process, there should be additional regulatory motivation for its inclusion.

The case study selected for this dissertation revolves around designing disease self-management technology for the diabetes population. Diabetes is becoming an increasingly prevalent chronic disease, with an estimated 415 million individuals diagnosed worldwide (Harding et al., 2019). Diabetes is responsible for substantial healthcare-related expenses in the United States, costing an estimated $245 billion in treatment and loss of productivity in 2012 (Menke et al., 2015). Non-adherence to diabetes disease self-management is a significant contributor to these costs and a serious problem in treating patients (Asche et al., 2011). Self-management of diabetes is highly dependent on supportive technology, for example glucometers and personal health record systems (Knisely & Vaughn-Cooke, 2020b). Therefore, methods for improving self-management of this disease are in high demand.

3.1 Background

This section presents additional background on patient-facing medical device use, and the challenges associated with engaging patient populations.

3.1.1 Patient-facing Medical Devices and User Error

Patient-facing medical devices and self-management technology have been shown to improve patient health outcomes, improve adherence to self-management, and decrease use of medical services (Asche et al., 2011; Greenwood et al., 2017; Pérez-Jover et al., 2019). Self-management technology is considered one the of best options for long-term patient care by helping patients take responsibility of their own health (Alessa et al., 2019). Despite these benefits, user error and issues with acceptance are common (Kannry et al., 2012; Silva et al., 2019). Commonly cited barriers to appropriate use include poor technological competence (Lyles et al., 2012; Pritchard & Nicholls, 2014), the ability of a patient to interact with novel technology, and poor health literacy (Mayberry et al., 2011; Shan et al., 2019), the ability of the patient to understand and apply information regarding their health. Specific types of user error experienced by patients range from difficulties following intended operating procedures, issues receiving device feedback, and difficulties performing physical interaction with the device (Knisely, Levine, et al., 2020).

There are many examples of usability issues occurring with marketed patient-facing products. One study evaluated three blood pressure monitors and found several issues related to equitable use (Cifter, 2017). Fung et al. (2015) discovered that individuals with physical or sensory impairments may experience interaction difficulties with positive airway pressure devices for treatment of sleep apnea.

Agnisarman et al. (2017) evaluated several home-based telemedicine software platforms and discovered significant differences in usability, user errors, and cognitive workload in participants interacting with the software.

Though these concerns remain common, patients have little influence on the design and evaluation of medical devices (Czaja et al., 2015; Ng et al., 2016). Incorporating feedback from patients during the medical device design process can help to mitigate downstream usability issues, consequences to patient safety, and prevent expensive recalls and litigations (Johansen, 2018). The US Food and Drug Administration requires that medical device manufacturers minimize unsafe device use by performing extensive human factors testing. While performing formative design validation, designers should evaluate device use with participants who are "representative of the range of characteristics within their user group," where each group should "perform different tasks or will have different knowledge, experience or expertise that could affect their interactions with elements of the user interface" (Food and Drug Administration, 2016). Given this, medical device manufacturers should strive to include patient populations in the design process, however, they face many challenges when engaging with these populations.

3.1.2 Engaging Patient Populations

Engaging with patients during the design process is often avoided due to perceived costs and delays in product development. The medical device industry is rapidly changing, and some manufacturers perceive formal methods of user engagement as detrimental to maintaining a competitive pace (Liao et al., 2015; Roma & de Vilhena Garcia, 2020). Organizational culture can prohibit engagement with users by observing preconceptions and discouraging the practice all together. Additionally, design teams may not have access to the user population. Further, when design teams do have access to the intended user population, they often lack the human factors experience and tools required for effective user engagement and device customization (Ozcelik et al., 2011).

Patient populations deviate significantly from the general population regarding design usability requirements (Czaja et al., 2015). This makes it even more important to specifically target patients when evaluating device design. In addition, the uniqueness of the patient population introduces new recruitment challenges. The majority of patient-facing medical device users have at least one chronic disease and are disproportionally represented as vulnerable and minority populations. This may include the disabled, racial and ethnic minorities, low socio-economic status, the elderly, and those who live in rural areas (UyBico et al., 2007). These populations come with unique recruitment challenges (McLaughlin et al., 2020). The elderly population, for example, has been reported as having general mistrust in institutions, transportation limitations, physical and cognitive impairment, apathy towards research participation, and medical and health related fears, all of which make access difficult (McHenry et al., 2015).

3.2 Objectives

The diabetes population is a safety-critical population, motivating the need for analysis of end-user capability in early-stage design (*upstream neutrality*) to identify necessary design requirements to mitigate user error. Further, this population is heterogeneous. Individuals with diabetes have varied cognitive skills required for device interaction (e.g., health literacy, technological competency), and have a high incidence of comorbidities (e.g., arthritis, glaucoma, neuropathy) that can influence product use (Showell, 2017; Trief et al., 2013; Weppner et al., 2010). This population is also non-general, motivating the need for currently non-existent human performance data. In each of the following chapters, a piece of the dissertation methodology will be introduced generally. Guidance will be provided for applying the method to any heterogeneous population. Then, the method piece will be demonstrated specifically for the diabetes population, with the goal to improve the design of diabetes self-management technology.

Chapter 4: Leveraging Physician Expertise and National Population Data to Model Heterogeneous Population User Groups

4.1 Introduction

In this chapter, a process for defining task-specific, performance-driven user groups utilizing expert input is proposed. Throughout this dissertation, experts are considered individuals who have significant experience or education with respect to human performance of routine human activities. It is proposed that internal medicine physicians satisfy these criteria. This medical specialty was targeted because they observe humans of all capabilities interacting with products (e.g., using medical devices) routinely and, as generalists, care for patients with diseases common among all the task domains included. They routinely assess and attempt to predict a patients' ability to perform tasks. Internal medicine physician expertise is not isolated to a specific sub-system of the human body and they therefore evaluate a broad range of task performance characteristics.

Figure 2 summarizes the process introduced in this chapter. First, an existing taxonomy of physical and cognitive tasks is translated into standardized tasks optimized for physician judgment (P&C Physician Judgment Tasks). Next, a process for using input from domain experts (physicians) to identify human characteristics relevant to the performance of these tasks is introduced. Then, a novel approach for mapping user characteristics to existing population data to be used as input for a cluster analysis is demonstrated. National population data was acquired from the National Health and Nutrition Examination Survey (NHANES) dataset (Centers for

Disease Control and Prevention, 2019). This data serves as input to statistical clustering to define user groups. This process can be replicated for any user population but is demonstrated on the diabetes population case study.

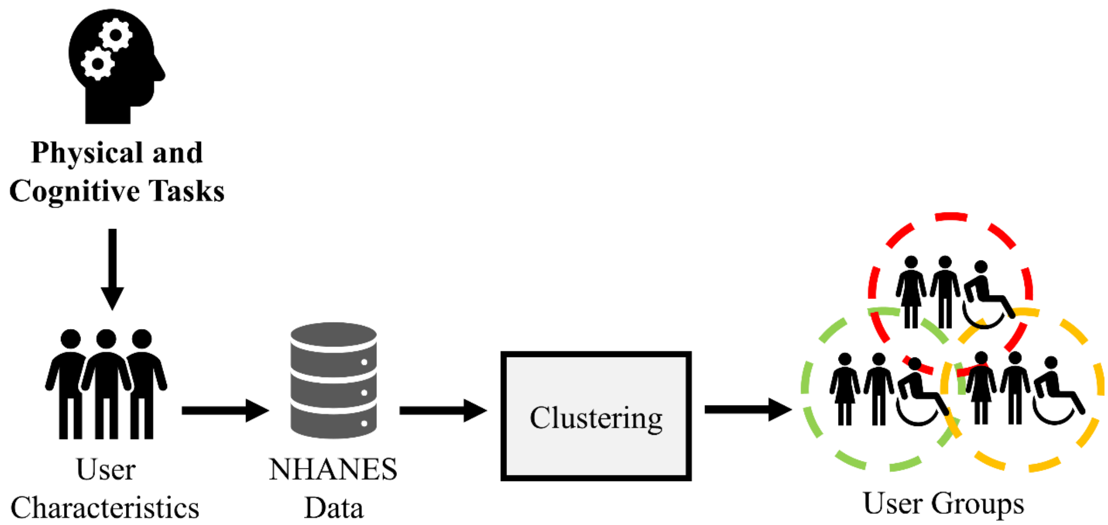


Figure 2: Summary of Chapter 4.

4.2 Methodology

This section describes the process for defining task-specific user groups. The steps for this methodology are summarized as follows:

1. Generate standardized physical and cognitive interaction tasks optimized for physician judgment (P&C Physician Judgment Tasks). E.g., Fine motor movement
2. Map tasks to relevant user characteristics. E.g., Disease history
3. Map user characteristics to NHANES variables. E.g., Presence of arthritis
4. Statistically cluster NHANES subjects to generate product user groups.

4.2.1 Generating Tasks for Physician Judgment

Human performance is highly dependent on the given context, which makes reusing human performance data across products difficult. In this work, standardized tasks required for product interaction were defined such that human performance data could be collected and applied across products for a given population. While making tasks general does create some uncertainty regarding accuracy, it extends the usefulness of collected data and facilitates future validation. Further, these tasks were defined to facilitate physician judgment on task performance, referred to as P&C Physician Judgment Tasks.

P&C Physician Judgment Tasks were derived from two existing taxonomies, including Bloom's taxonomy of the cognitive domain (Bloom, 1956) and Harrow's taxonomy of the psychomotor domain (Harrow, 1972). While originally intended for and most commonly applied to evaluating educational objectives (Crompton et al., 2019; Verenna et al., 2018), Bloom's and Harrow's taxonomies were repurposed to describe product interaction from a human factors perspective. Both taxonomies have been subject to validating efforts as tools for classifying the complexity of human activities (Hamid et al., 2012; Knisely, Joyner, et al., 2020; Lalwani & Agrawal, 2018; Phillips et al., 2013; Roberts, 1976; Soozandehfar & Adeli, 2016). Tables 2-3 contains Bloom's and Harrow's Taxonomy, tasks ordered by increasing complexity.

Table 2: Bloom's Taxonomy listed in order of increasing cognitive complexity.

Taxonomy Level	**Description**
Knowledge	Recall of specific facts or ideas.
Comprehension	Understanding and interpreting facts and ideas.
Application	The use of prior knowledge in novel situations.
Analysis	Decomposing a system into its composite parts and examining those parts.
Synthesis	Combining independent elements to form a new system.
Evaluation	Judging the value of a system based on evidence and certain criteria.

Table 3: Harrow's Taxonomy listed in order of increasing psychomotor complexity.

Taxonomy Level	Description
Reflexive Movements	Involuntary movements evoked in response to some stimuli.
Fundamental Movements	Basic movement patterns which build on reflexive movements and include acts such as reaching, grasping, and walking.
Perceptual Abilities	Ability to receive information about oneself and the world via one of several sensory systems (vision, hearing, etc.).
Physical Abilities	The functional characteristics of the body which govern the efficiency of skills in the psychomotor domain.
Skilled Movements	Complex movement skills that require learning.
Non-Discursive Movements	Learned movements used for communication.

The researchers used their combined expertise in human factors and physician judgment to decompose and translate levels of Bloom's and Harrow's taxonomy into tasks that were: a) typical of product interaction and b) conformed with physician understanding of what constitutes a distinct task. The objective was to provide a model for describing product interaction such that physicians could make quantitative judgments on task performance. The granularity of tasks was evaluated based on trade-offs between how specific they could be applied verses the complexity of the resulting analysis. More specific tasks can be used to describe device interaction more precisely, but would create a larger decomposition, and increase the number tasks to be analyzed in subsequent steps. Refining these tasks further for specific applications may be required, as not all tasks may be relevant for every product. Table 4 demonstrates this methodological development.

Table 4: Methodological development (left to right) for translating taxonomy tasks into tasks tailored for physician judgment with specific examples.

		P&C Taxonomy Tasks	P&C Physician Judgment Tasks	Examples
Intended Use	Scope (Generalizability)	Product interaction	Product interaction	-
	User Population	All	All	-
	Terminology tailored for:	Human factors	Physician judgment	-
Harrow's Taxonomy		Reflexive Movement	-	-
		Fundamental Movement or Physical Abilities*	Gross Upper-body Movement	Pushing, pulling, holding
			Gross Lower-body Movement	Kicking, stomping, squatting
			Locomotor Movement	Walking, running, crawling
			Fine Motor Movement	Pressing, twisting, grasping
			Speech Production	Talking
		Perceptual Abilities	Visual Discrimination	Seeing
			Auditory Discrimination	Hearing
			Tactile Discrimination	Feeling
		Skilled Movements	-	-
		Non-discursive Communication	-	
Blooms Taxonomy		Knowledge	Applying Existing Knowledge	Recalling knowledge, classifying knowledge, executing procedures
		Comprehension		
		Application		
		Analysis	Problem-solving and Decision-making	Comparing, contrasting, assembling, integrating, judging, critiquing
		Synthesis		
		Evaluation		

*Physical Abilities is typically listed below perceptual abilities, however several physician judgment tasks could conceivably be group as Fundamental Movement or Physical Abilities, so they were grouped.

Tasks listed under Harrow's taxonomy are psychomotor, requiring motor and neuromuscular control (Harrow, 1972). Note that while Physical Abilities appears below Perceptual Abilities in Harrow's taxonomy, Fundamental Movements and Physical Abilities were grouped in Table 4 because several of the generated P&C

Physician Judgment Tasks could conceivably belong to either. Only visual, auditory, and tactile discrimination were included for Perceptual Abilities because they represent the typical sensory modalities needed for product interaction. No generated tasks fell under Reflexive Movement, Skilled Movements, or Non-discursive Movements. Reflexive Movement were determined to be too fundamental to be useful to describe a complete product interaction. Skilled Movements are product specific by their very nature and can therefore not be generalized for typical product interaction. Non-discursive communication describes movements used for communication with another human. While gesturing is used for communication with some autonomous systems, it was determined to be an edge case and was not considered.

Tasks listed under Bloom's taxonomy are those that require mental processing. The tasks included were primarily unobservable actions. Tasks in Bloom's taxonomy are presented in order of increasing amounts of conscious control required for execution. The individual levels of Bloom's taxonomy were identified as too granular to facilitate physician judgment and were categorized into two groups. These categories include Applying Existing Knowledge and Problem-solving and Decision-making. Applying Existing Knowledge requires using existing knowledge in a routine way. Problem-solving and Decision-making tasks require the creation of new knowledge or applying old knowledge to a new situation (Krathwohl, 2002), which is considered a more cognitively complex activity. The main distinction between these cognitive tasks is that an existing rule (conditional statement) is used to apply previously acquired knowledge for Applying Existing Knowledge, whereas a new rule must be created to perform Problem-solving and Decision-making tasks.

4.2.2 Identifying Relevant User Characteristics via Expert Input

The next step in the methodology is to identify performance-driving user characteristics for each P&C Physician Judgment Task in Table 4. To help identify these variables, a survey was devised for domain experts to rank the importance of certain characteristics to the performance of the tasks. Internal medicine physicians from the University of Maryland Medical Center were targeted for recruitment.

Three high-level tasks were included in the survey: physical, sensory and perception, and cognition. For each, experts were given a list of user characteristics and asked to rank them based on the order they would consider them when evaluating the ability of an individual to perform the task. Experts were required to rank at least one user characteristic per task. The survey was developed and administered using Qualtrics. All internal medicine physicians affiliated with the University of Maryland Medical Center were eligible to participate. The questions were presented as follows:

1. If asked to predict the ability of a population to perform tasks requiring physical effort, what information would you consider? In what order would you consider it?
2. If asked to predict the ability of a population to perform tasks requiring sensation and perception, what information would you consider? In what order would you consider it?

3. If asked to predict the ability of a population to perform tasks requiring problem-solving and decision-making (cognition), what information would you consider? In what order would you consider it?

The survey was administered to expert physicians with no direct incentives. Therefore, it was critical to limit the number of questions and survey time to ensure adequate recruitment levels and participation. The subject of each question (physical effort, sensation and perception, problem-solving and decision-making) were identified as the lowest level of specificity that could be presented to physicians and still elicit meaningful judgments, while not compromising the data collection goals.

User characteristics to be included were generated via an iterative process of requirements elicitation and refinement. Existing literature, and the co-authors expertise in medicine and product interaction, were used to create a preliminary list. The focus of inclusion was user characteristics with a direct potential influence on the performance of tasks. The characteristics and justification for inclusion are shown in Table 5.

Table 5: User characteristics included in the medical expert survey.

Patient Characteristic	Justification for Inclusion
Age	Age is associated with decline of physical (Seidler et al., 2010; Senefeld et al., 2017), sensory (Rudman et al., 2016; P. Wu et al., 2020), and cognitive abilities (Meng et al., 2017).
Socioeconomic Status	Socioeconomic status has been shown to be associated with various health outcomes (Blackwell et al., 2014; Präg et al., 2016).
Physical Independence	Level of physical independence has been shown to be associated with physical activity (Marques et al., 2014) and cognitive task performance (Sobol et al., 2016).
Decision-making	Decision-making, attention, and memory skills are integral elements of cognition (Guilera et al., 2020).
Attention	
Memory	
Substance Abuse	Substance abuse has been linked to impairment of physical, sensory, and cognitive abilities (Barnes, 2014; Harvey et al., 2018; Toplak et al., 2010).
Exercise	Exercise has been linked to physical activity (Liubicich et al., 2012; Rejeski et al., 2010).
Psychiatric Disorder	Psychiatric disorders have been linked to impaired cognition (Toplak et al., 2010).
Disease History	Many chronic diseases have been associated with poor human performance (Fung et al., 2015; Showell, 2017).
Disease Severity	
Details of Task	Individual as well as contextual factors determine task performance.
Health Literacy	User health literacy has been associated with perceived medical device usability (Chaniaud et al., 2020).
Other (please elaborate in text box):	-

Each participant was asked to rank the characteristics that would be considered to predict outcomes for the aforementioned tasks, along with the order that they would consider this information. Rankings were evaluated using the Borda count (Emerson, 2013), where characteristics ranked 1 received n points, characteristics ranked 2 received n – 1, and so on, where n is the total number of options. User characteristics with the highest counts summed across experts are used in next stages to identify relevant user data.

4.2.3 Mapping Tasks and User Data

Following identification of task-relevant user characteristics, user characteristics can be quantified for a given population by mapping them to variables in existing data. Several publicly available databases exist with data on human health, capabilities, etc. Examples include NHANES (Centers for Disease Control and Prevention, 2019), the National Health Interview Survey (Centers for Disease Control and Prevention, 2018), and the US Army Anthropometric Survey (ANSUR) (Gordon et al., 1989). In this work, NHANES variables are linked to user characteristics. NHANES is a longitudinal survey used to monitor the health and wellbeing of United States citizens. Demographic, socioeconomic, dietary, and health-related questions are included. This data is well suited to define user groups because of the many physical and cognitive health characteristics included and its coverage of the specific characteristics identified in the physician survey.

In the previous section, a process to identify user characteristics relevant to task performance was described. Following this, corresponding NHANES variables are mapped to each user characteristic. These variables can be used to cluster subjects into task-specific user groups. Using NHANES data ensures that the composition of user groups reflects the actual population and eliminates the risk of customizing a product for users who are not well represented. NHANES is released in yearly installments, and not all variables are consistent from year to year. In this work, NHANES 2017-2018 is used.

For each variable, justification was sought in literature for the mapping. Specific diseases were justified using the WHO ICF core sets. ICF core sets link standardized terminology for human functioning and health to specific disease categories (Selb et al., 2015). If a standard term analogous to the task being assessed was present in an ICF disease core set, then it was assumed the corresponding NHANES disease variable is relevant to the task. The steps of this justification process are summarized as follows:

1. Find an NHANES variable for a specific disease. (e.g., Arthritis)
2. Hypothesize link between NHANES variable and cognitive or psychomotor task. (e.g., NHANES Variable: Arthritis → Task: Fine Motor Movement)
3. Search ICF for standard terminology for human functioning analogous to the task. (e.g., Task: Fine Motor Movement → ICF Term: Fine Hand Use)
4. Find corresponding ICF core set for disease and locate ICF core set term (e.g. ICF course set: post-acute musculoskeletal disease (Scheuringer et al., 2005) → ICF Term: Fine Hand Use)

Following selection and justification of NHANES variables, data can be retrieved from the NHANES website. The data is split into multiple subject specific files that must be retrieved separately. Participants are labeled with a unique "Respondent Sequence Number" that can be used to link data together. If a specific chronic disease population is of interest, and the appropriate NHANES variable exists, participants can be filtered based on having the chronic disease.

4.2.4 Clustering User Data

To define user groups, the data is statistically clustered. It is common for several clustering algorithms to be benchmarked against one another when clustering data. Three clustering algorithms were selected for this work, including gaussian mixture models (GMM), partitioning around medoids (PAM), and hierarchical clustering (HC). This selection of clustering algorithms represents three of the common classes of clustering algorithms used for mixed-data types – model-based (GMM), partitional (PAM), and hierarchal (HC) (Ahmad & Khan, 2019). Using multiple clustering algorithms can help illuminate general cluster-based trends in the data and can produce a larger variety of candidate solutions to evaluate.

GMMs are a model-based clustering algorithm that assume data exist as several sub-populations that follow gaussian distributions (Ahmad & Khan, 2019). Distribution means and variances are fit to the data using the expectation-maximization algorithm, allowing the probability of membership for each cluster to be calculated for each data point. Thus, data points are given mixed or "soft" assignments to clusters. Further, because the data was mixed (continuous, ordinal, and binary), the R package clustMD is used because it is formulated to accept mixed-data types (McParland & Gormley, 2015). While GMM clustering does have a relatively higher time complexity than simpler clustering algorithms (D. Xu & Tian, 2015), the soft group assignment used with GMMs fits with the logic for user group membership. The variables used to define clusters explain a portion of the variance in task performance that would be observed in practice. Certainly, there are other user characteristics not included in NHANES that could provide additional information, as well as other variables regarding the context of the task being performed that are unknowable. With GMM, this uncertainty is incorporated into the clusters by assigning probabilities for group membership.

PAM is a more robust version of the k-means algorithm and is better suited for mixed-data type clustering. While k-means fits clusters using Euclidean distance and identifies cluster centers using cluster means (centroid), PAM accepts arbitrary distance metrics and restricts cluster centers to be actual members of the data (referred to as a medoid) (Schubert & Rousseeuw, 2019). HC is a flexible clustering method that assigns each data point to a unique cluster, and iteratively merges clusters based on proximity given a selected distance metric (Murtagh & Contreras, 2017). For both algorithms, the Gower distance is utilized. Gower distance accepts continuous, ordinal, and categorical variables and produces an aggregate distance measure (Podani, 1999). Both PAM and HC are less complex clustering algorithms than GMM and serve as model benchmarks.

To evaluate the clustering algorithms and the number of clusters, a mixed internal and external validation approach is taken. The advantage of using both internal and external validation is that it utilizes both prior knowledge about the subject matter and new information intrinsic to the data. This prevents overreliance on preconceptions about the structure of the data but also ensures that generated clusters are meaningful (Gajawada & Toshniwal, 2012). For internal validation, several commonly used metrics were identified. The metrics selected include the silhouette index, Calinski-Harabasz (CH) index, connectivity, and Bayesian Information Criterion (BIC). Silhouette index and CH index both measure cluster compactness, a measure of intra-cluster variance, along with cluster separation (Brock et al., 2008; Caliński & Harabasz, 1974). While these metrics measure similar qualities, they demonstrate different performance given various properties of the data (e.g., noisy data, cluster skewedness) (Y. Liu et al., 2010). Both should be maximized. Connectivity is a measure of the connectedness between clusters and should be minimized (Handl et al., 2005). BIC selects the number of clusters that maximizes model likelihood (goodness of fit) while penalizing model complexity (number of parameters in each model) (McParland & Gormley, 2015). BIC relies on model likelihood estimations and is therefore only available for GMM. Clustering algorithms fit clusters by optimizing different objective functions and may be biased towards certain validation metrics. Therefore, internal validation metrics are only used to evaluate the number of clusters within algorithms, not across.

For external validation, dominant cluster characteristics are extracted, summarized, and evaluated subjectively by the researchers. Each cluster are qualitatively assigned a relative risk category, where the highest and lowest risk clusters corresponding to the highest and lowest risk categories. Clusters can then be evaluated based on qualitative separation, conformity with researcher expectations given their medical backgrounds, and for anticipated usefulness for product personalization. Clusters that are qualitatively separated are desirable for the customization task because it helps to justify specifically tailored design solutions. Quantitative cluster performance in isolation does not guarantee meaningfully distinct clusters in practice.

4.3 Case Study Application

The above methodology was applied to the diabetes self-management case study. This section discusses how the methodology was specifically tailored for this application.

4.3.1 Case Study Tasks

Tasks were evaluated for inclusion based on minimizing the tasks required to characterize the population, which will become critical when expert elicitation is used to quantify task performance (Chap. 5) and recruitment efforts could be hampered by a lengthy process. This required elimination of tasks that were not directly relevant to diabetes self-management. The final list of tasks to be included are summarized in Table 6. The physical tasks not included were determined to be irrelevant for most medical device interaction related to diabetes. Finally, while relevant, tactile discrimination and speech production were eliminated for this particular case study due to the limited relevant data in the NHANES database. The included tasks were used to drive the identification of relevant user characteristics by experts, which drove data acquisition and clustering.

Table 6: Methodological development (left to right) for translating taxonomy tasks into diabetes case study tasks.

		P&C Taxonomy Tasks	**P&C Physician Judgment Tasks**	**P&C Case Study Tasks**
Intended Use	Scope	Product interaction	Product interaction	Medical device interaction
	User Population	All	All	Diabetes population
	Terminology tailored for:	Human factors	Physician judgment	Physician judgment
Harrow's Taxonomy		Reflexive Movement	-	-
		Fundamental Movement or Physical Abilities	Gross Upper-body Movement	Gross Upper-body Movement
			Gross Lower-body Movement	†
			Locomotor Movement	†
			Fine Motor Movement	Fine Motor Movement
			Speech Production	*
		Perceptual Abilities	Visual Discrimination	Visual Discrimination
			Auditory Discrimination	Auditory Discrimination
			Tactile Discrimination	*
		Skilled Movements	-	-
		Non-discursive Communication	-	-
Blooms Taxonomy		Knowledge	Applying Existing Knowledge	Applying Existing Knowledge
		Comprehension		
		Application		
		Analysis	Problem-solving and Decision-making	Problem-solving and Decision-making
		Synthesis		
		Evaluation		

* Removed due to lack of data, † Removed due to lack of relevance.

4.3.2 User Clustering

Data Preparation: Data retrieved from NHANES was filtered by subjects who reported having diabetes. Only participants who responded to all the questions for each task were kept in each dataset. Therefore, the samples per task was dependent on the NHANES variables included. Further, individuals younger there 18 were not included in the analysis. Sample sizes for cluster analyses ranged from 616 to 720. Exact sample sizes for each cluster analysis are presented in the results. Continuous variables were scaled by subtracting the mean and dividing by the standard deviation. Dichotomous variables were translated into binary values, where 1 indicated the presence of that variable (e.g., 0 = No arthritis, 1 = Has arthritis). Ordinal variables were translated into integers corresponding to the number of levels. For clustering algorithms not explicitly designed to handle mixed data (PAM, HC), the pairwise distance matrix was calculated for all datapoints using Gower distance prior to clustering.

Clustering Details: Cluster numbers and clustering algorithms were evaluated internally and externally as detailed in Chapter 4.2.4. Cluster quantity was limited to 4 to provide a feasible range of user groups for the subsequent expert elicitation. ClustMD includes several covariance structure models that can be used that allow varying degrees of GMM complexity in terms of cluster volume and orientation (McParland & Gormley, 2015). Parameters that control cluster shape and orientation are either set as the identity matrix, constrained across clusters, or totally unconstrained. All six covariance models were tested for 2, 3, and 4 clusters. The models with the best performance (BIC) for each cluster count were selected and subjected to internal and external validation.

4.4 Results

Discussed in this section are the results from the survey to identify relevant user characteristics for task performance, the results of the user characteristic – NHANES variable mapping, and results of the cluster analysis.

4.4.1 Expert Identified User Characteristics Survey

Figure 3 displays the Borda counts for the three survey questions on user characteristics and task performance.

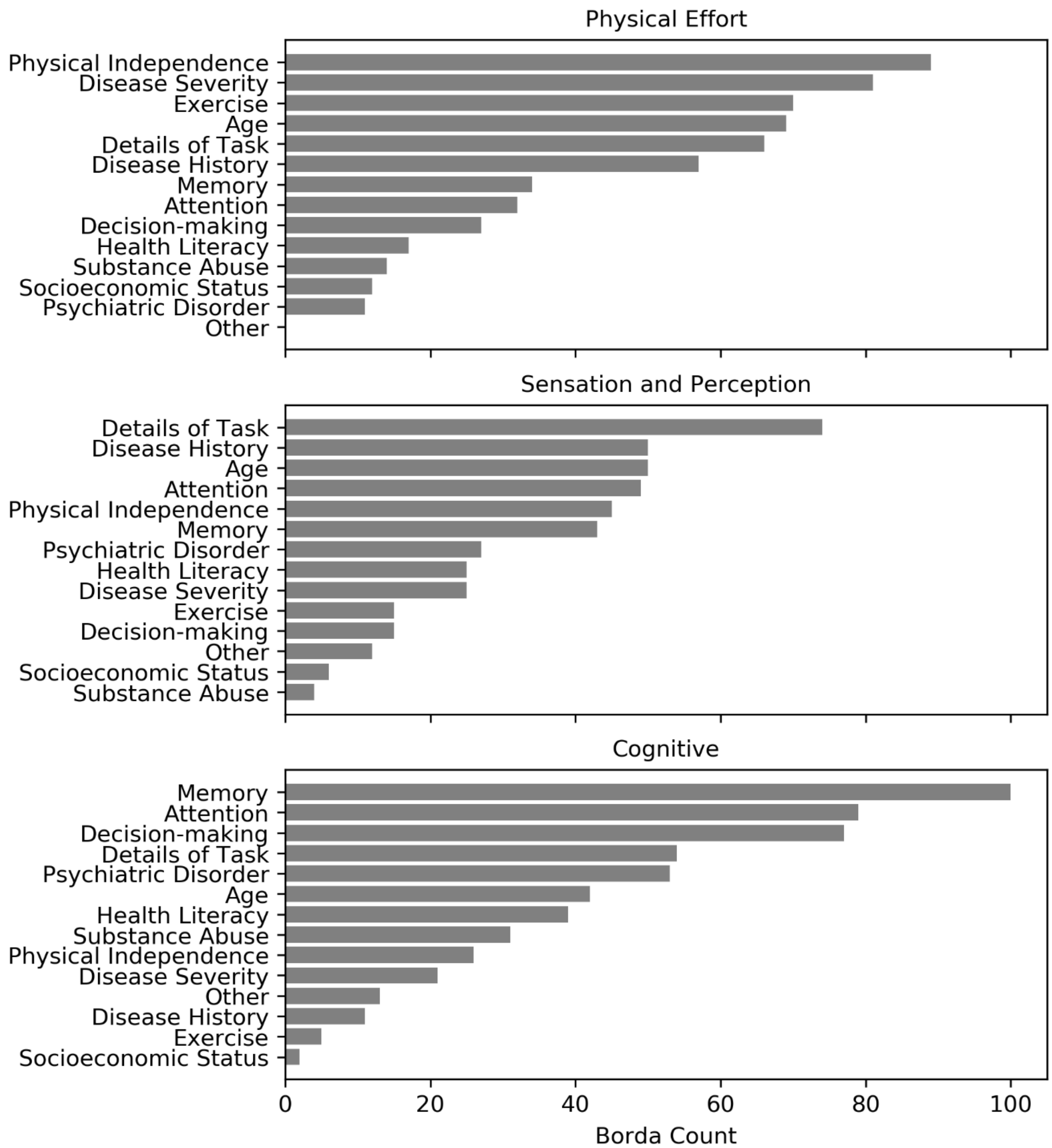


Figure 3: Borda count for each user characteristic for the three high-level tasks included.

For each survey task, the top 6 characteristics were selected for use in identifying relevant variables from the NHANES database. The task-characteristic mappings are as follows:

- **Gross Upper-body Movement (Physical Effort):** Physical Independence, Disease Severity, Exercise, Age, Details of Task, Disease History
- **Fine Motor Movement (Physical Effort):** Physical Independence, Disease Severity, Exercise, Age, Details of Task, Disease History
- **Visual Discrimination (Sensory and Perception):** Details of Task, Disease History, Age
- **Auditory Discrimination (Sensory and Perception):** Details of Task, Disease History, Age
- **Applying Existing Knowledge (Cognitive):** Memory, Attention, Decision-making, Details of Task, Psychiatric Disorder, Age
- **Problem-solving and Decision-making (Cognitive):** Memory, Attention, Decision-making, Details of Task, Psychiatric Disorder, Age

For visual and auditory discrimination, only the top 3 characteristics were selected as directly relevant to sensory and perception. Characteristics 4-6 were considered but determined to be ultimately not relevant enough for inclusion.

4.4.2 User Characteristic – NHANES Variable Mapping

The NHANES database was reviewed for variables relevant to each user characteristic. Table 7 contains the NHANES variables mapped to characteristics for “gross upper-body movement”. Tables for the other tasks can be seen in Appendix A. The format of the NHANES variable and justification for inclusion are also included. “Details of task”, while determined to be important, was not included because it was not represented in NHANES as it is a contextual characteristic as opposed to a patient characteristic. In some cases, where disease severity was available, disease history and disease severity were converted into a single, ordinal variable.

Table 7: User characteristic - NHANES variable mapping for "gross upper-body movement."

Expert Survey Characteristic	NHANES Variable	Format	Justification
Age	Age	Continuous	Age and decreased muscular strength are associated (Senefeld et al., 2017).
Physical Independence	Reported difficulty dressing or bathing	Binary – Yes or No	Activity is a specific case of “gross upper-body movement.”
Physical Independence	Reported difficulty reaching up	Ordinal – No difficulty, Some difficulty, Much difficulty, Unable to do, Does not do	Activity is a specific case of “gross upper-body movement.”
Physical Independence	Reported difficulty moving large objects	Ordinal – No difficulty, Some difficulty, Much difficulty, Unable to do, Does not do	Activity is a specific case of “gross upper-body movement.”
Disease History	Reported having: - Arthritis - Gout - Bone/joint injury - Neck and back problems	Binary – Yes or No	“Hand and arm use” linked with post-acute musculoskeletal disease ICF core set (Scheuringer et al., 2005).
Disease History	Reported having: - Congestive heart failure - Angina/angina pectoris	Binary – Yes or No	“Hand and arm use” linked with cardiopulmonary post-acute ICF core set (Wildner et al., 2005).
Disease History	Reported having coronary heart disease	Binary – Yes or No	“Lifting and carrying objects” linked with ischaemic heart disease ICF core set (Cieza et al., 2004).

Table 7 (continued)

Disease History	Reported having: - Asthma - Emphysema - Chronic Bronchitis - COPD	Binary – Yes or No	"Lifting and carrying objects" linked with obstructive pulmonary disease ICF core set (Stucki, Stoll, et al., 2004).
Disease History	Reported having obesity	Binary – Yes or No	"Lifting and carrying objects" linked with obesity ICF core set (Stucki, Daansen, et al., 2004).
Disease History	Reported having a stroke	Binary – Yes or No	"Hand and arm use" linked with stroke ICF core set (Geyh et al., 2004).
Exercise	Reported physical activity at work	Ordinal – None, Moderate, Vigorous	Physical activity is associated with upper-body mobility (Rejeski et al., 2010).
Exercise	Reported physical activity recreationally	Ordinal – None, Moderate, Vigorous	Physical activity is associated with upper-body mobility (Rejeski et al., 2010).

4.4.3 User Group Cluster Analysis

Using the prior variables, data was statistically clustered using each clustering algorithm. Cognitive task cluster results ("applying existing knowledge" and "problem-solving and decision-making") were evaluated together because the NHANES variables identified for both were identical, and therefore clusters were not differentiated. GMM BIC values per cluster per task are shown in Appendix B. Note that in the R packages used, BIC is formulated such that the maximum value is sought, while in many other cases it is formulated such that the minimum is sought.

Tables 8-10 contain cluster internal validation metrics and subjective risk level categories for "gross upper-body movement", "fine motor movement", and cognitive tasks. Summaries of dominant characteristics for each cluster solution that were translated into subjective risk-levels are displayed in Appendix C.

Table 8: Internal and external validation criteria for "gross upper-body movement" clusters.

Clustering Algorithm	**PAM**			**HC**			**GMM**		
Cluster Count	2	3	4	2	3	4	2	3†	4
Silhouette Index	0.184*	0.135	0.141	0.173	0.198*	0.163	0.167*	0.124	0.110
Connectivity	127.7*	310.2	254.8	178.2*	232.1	254.0	341.3*	469.1	607.9
CH Index	161.1*	123.5	117.2	71.3	76.2*	69.3	126.2*	104.6	85.9
BIC (× 10^4)	-	-	-	-	-	-	-1.65	-1.64*	-1.65
C1 RL	ML	L	L	M	ML	ML	ML	L	ML
C2 RL	MH	ML	ML	MH	MH	M	MH	M	ML
C3 RL	-	MH	M	-	MH	MH	-	H	ML
C4 RL	-	-	MH	-	-	MH	-	-	H

C# = Cluster #, RL = Risk-level, L = Low, ML = Moderately Low, M = Moderate, MH = Moderately High, H = High, *Highest scoring cluster count for each metric and each algorithm, †Selected cluster solution.

Table 9: Internal and external validation criteria for "fine motor movement" clusters.

Clustering Algorithm	**PAM**			**HC**			**GMM**		
Cluster Count	2	3	4	2	3	4	2†	3	4
Silhouette Index	0.314*	0.287	0.291	0.158	0.144	0.194*	0.235*	0.104	0.138
Connectivity	14.5*	116	146	48.0*	49.5	68.5	224*	315	357
CH Index	298*	209	194	74.7	40.0	85.6*	127.3*	71.0	79.6
BIC (× 10^3)	-	-	-	-	-	-	-9.82*	-9.85	-9.88
C1 RL	ML	ML	ML	ML	ML	ML	M	ML	ML
C2 RL	M	ML	ML	M	M	M	MH	M	M
C3 RL	-	ML	M	-	H	MH	-	H	MH
C4 RL	-	-	MH	-	-	H	-	-	MH

C# = Cluster #, RL = Risk-level, L = Low, ML = Moderately Low, M = Moderate, MH = Moderately High, H = High, *Highest scoring cluster count for each metric and each algorithm, †Selected cluster solution.

Table 10: Internal and external validation criteria for cognitive task clusters.

Clustering Algorithm	**PAM**			**HC**			**GMM**		
Cluster Count	2	3	4	2	3	4	2	3†	4
Silhouette Index	0.355*	0.313	0.265	0.309	0.217	0.319*	0.295	0.303*	0.301
Connectivity	136	95.2*	182	49.1*	67.6	79.9	40.8*	106	133
CH Index	362*	357	294	93.5	71.5	173*	300	316*	280
BIC (× 10^4)	-	-	-	-	-	-	-1.23	-1.15*	-1.19
C1 RL	ML	L	L	M	ML	ML	L	L	L
C2 RL	MH	M	M	M	M	ML	M	M	M
C3 RL	-	MH	M	-	M	M	-	H	MH
C4 RL	-	-	MH	-	-	MH	-	-	H

C# = Cluster #, RL = Risk-level, L = Low, ML = Moderately Low, M = Moderate, MH = Moderately High, H = High, *Highest scoring cluster count for each metric and each algorithm, †Selected cluster solution.

For "gross upper-body movement" and cognitive tasks, the 3-cluster solutions with GMM had at least one internal validation metric that was optimal and produced the most qualitatively demarcated clusters and were thus selected as optimal. For fine motor movement, HC with 4 clusters produced the best qualitative separation, however cluster 4 only contained one individual, so these results were eliminated from contention. Remaining options with optimal validation metrics were 2-cluster options. While equally qualitatively separated, GMM was selected over PAM or HC because the model produced higher risk groups. The researchers determined that taking a more conservative approach to represent risk was preferable from a perspective of maximizing safety via personalization.

For visual and auditory discrimination, only age and one additional variable were identified in NHANES for inclusion. For both, the additional variable was subject self-reported ability to see and hear. Initial attempts to cluster these variables produced clusters that could not be practically differentiated for the purposes of device customization. As such, statistical clustering was abandoned for these tasks in favor of subjective clustering. Clusters were manually formed by grouping the self-reported task performance variables. For visual discrimination, this was a binary variable, and therefore two clusters resulted. For auditory discrimination, the variable contained six levels of hearing quality. Given the imposed 2-4 cluster constraint, the authors determined grouping the levels into twos (i.e., three clusters) produced the most meaningful clusters for differentiating patients by performance. "Wears a hearing device" is presented with the auditory discrimination clusters. This was not used in the clustering and was only included to verify “reported hearing quality without correction” was representative of day-to-day hearing for the individual.

Cluster results for the selected cluster solutions are shown in Tables 11-20. As referenced in Appendix A, NHANES variables are mostly binary or ordinal. For those identified as ordinal, reported levels are provided in the results tables. In addition to the data on individual cardio-pulmonary conditions, the aggregate number of cardio-pulmonary conditions per patient for gross upper-body movements is presented. Aggregate cardio-pulmonary conditions were calculated by summing the number of cardio-pulmonary diseases an individual had, including congestive heart failure, angina/angina pectoris, coronary heart disease, asthma, emphysema, chronic bronchitis, and COPD.

Highlighted values in the cluster results indicate majority and consideration of those factors as dominant for the cluster. For ordinal variables where a single level did not dominate (all levels contain <50% of individuals), a value is highlighted such that the majority of individuals in the cluster reported that level or worse. To support visualization, the dominant variables are highlighted based on their putative relationship with performance outcomes, with green indicating a positive relationship, red indicating a strong negative relationship, and orange indicating a moderate negative relationship.

Table 11: Individuals in clusters (n) and proportion of population for "gross upper-body movement."

	Cluster 1	**Cluster 2**	**Cluster 3**
n (%)	300 (41.7)	256 (35.6)	164 (22.8)

Table 12: GMM clusters for "gross upper-body movement" task. Values correspond to the variable in first column. Excluding age, this is the proportion of individuals in the cluster who reported that characteristic.

NHANES Variable	Level	Cluster 1	Cluster 2	Cluster 3
Median Age	-	68	65	70
Physical Independence				
Difficulty dressing and bathing (%)	-	0.333	25.8	33.5
Difficulty reaching up (%)	No difficulty	**98.0**	**41.8**	37.8
	Some difficulty	2.0	39.8	**43.3**
	Much difficulty	0	14.5	9.15
	Unable to do	0	3.13	7.32
	Does not do	0	0.780	2.44
Difficulty moving large objects (%)	No difficulty	**83.7**	14.8	8.54
	Some difficulty	15.0	**41.4**	29.9
	Much difficulty	1.33	19.5	**24.4**
	Unable to do	0	17.6	24.4
	Does not do	0	6.64	12.8
Disease History				
Arthritis (%)	-	39.3	**68.4**	**81.7**
Gout (%)	-	17.0	9.77	22.0
Bone/joint injury (%)	-	4.33	29.3	17.1
Neck and Back Problem (%)	-	7.30	**64.5**	**53.7**
Stroke (%)	-	6.67	14.1	24.4
Obesity (%)	-	50.0	**63.7**	**70.7**
Congestive heart failure (%)	-	6.67	0	48.2
Angina/angina pectoris (%)	-	7.00	1.56	29.2
Coronary heart disease (%)	-	10.7	1.17	45.7
Asthma (%)	-	6.67	19.5	37.2
Emphysema (%)	-	1.33	0.40	15.9
Chronic bronchitis (%)	-	3.33	10.2	40.9
COPD (%)	-	4.67	0.781	46.3
Total cardio-pulmonary conditions (%)	0	**71.3**	**70.3**	0
	1	19.3	25.8	22.0
	2	7.30	3.90	25.6
	3	1.70	0	**28.0**
	4+	0.300	0	24.4
Exercise				
Physical work activities (%)	None	**61.7**	**64.1**	**65.9**
	Moderate	21.3	20.7	19.5
	Vigorous	17.0	15.2	14.6
Physical recreational activities (%)	None	**59.3**	**74.2**	**84.8**
	Moderate	31.7	21.1	13.4
	Vigorous	9.00	4.70	1.80

Table 13: Individuals in clusters (n) and proportion of population for "fine motor movement" task.

	Cluster 1	Cluster 2
n (%)	536 (74.4)	184 (25.6)

Table 14: GMM clusters for "fine motor movement" task. Values correspond to the variable in first column. Excluding age, this is the proportion of individuals in the cluster who reported that characteristic.

NHANES Variable	Level	Cluster 1	Cluster 2
Median Age	-	66	72
Physical Independence			
Difficulty using fork, knife, cup (%)	No difficulty	**100**	**69.3**
	Some difficulty	0	30.6
	Much difficulty	0	4.89
	Unable to do	0	0.543
	Does not do	0	0.543
Difficulty grasp/holding small objects (%)	No difficulty	**89.1**	26.9
	Some difficulty	10.9	**73.5**
	Much difficulty	0.373	19.0
	Unable to do	0	4.35
	Does not do	0	0
Disease History			
Arthritis (%)	-	**50.4**	**85.3**
Gout (%)	-	15.5	15.8
Bone/joint injury (%)	-	12.5	26.6
Stroke (%)	-	6.72	32.6
Congestive heart failure (%)	-	5.60	37.5
Angina/angina pectoris (%)	-	5.60	23.4
Exercise			
Physical work activities (%)	None	**59.0**	**76.6**
	Moderate	22.0	16.8
	Vigorous	19.0	6.52
Physical recreational activities (%)	None	**65.9**	**83.7**
	Moderate	26.9	14.7
	Vigorous	7.28	1.63

Table 15: Individuals in clusters (n) and proportion of population for "applying existing knowledge" and "problem-solving and decision-making."

	Cluster 1	**Cluster 2**	**Cluster 3**
n (%)	270 (43.8)	247 (40.1)	99 (36.7)

Table 16: GMM clusters for "applying existing knowledge" and "problem-solving and decision-making". Values correspond to the variable in first column. Excluding age, this is the proportion of individuals in the cluster who reported that characteristic.

NHANES Variable	**Level**	**Cluster 1**	**Cluster 2**	**Cluster 3**
Median Age	-	69	67	62
Memory, Attention, and Decision-making Skills				
Problems managing money (%)	No difficulty	**91.1**	**84.6**	**57.6**
	Some difficulty	3.70	10.2	31.3
	Much difficulty	1.11	1.21	5.05
	Unable to do	0.740	0.810	2.02
	Does not do	3.33	3.24	4.04
Frequency feeling tired or low energy over a two-week period (%)	Not at all	**70.0**	36.8	12.1
	Several days	19.2	**42.5**	27.3
	More than half	5.56	10.1	**18.2**
	Nearly every day	5.19	10.5	42.4
Reports confusion/memory problems (%)	-	8.90	15.4	**55.6**
Reports serious difficulty concentrating, remembering, or making decisions (%)	-	7.04	14.2	**61.6**
Disease History				
Stroke (%)	-	8.52	14.7	20.2
Psychiatric Disorder				
Anxiety frequency (%)	Never	**53.0**	9.31	0
	Few times a year	35.9	**43.3**	1.01
	Monthly	8.52	20.7	5.05
	Weekly	1.11	17.4	19.2
	Daily	1.48	9.31	**74.7**
Anxiety severity (if reported) (%)	Mild	**65.6**	46.0	8.08
	Moderate	25.0	**45.1**	**50.5**
	Severe	9.40	8.90	41.4
Depression frequency (%)	Never	**100**	1.63	0
	Few times a year	0	**75.1**	5.05
	Monthly	0	16.7	6.06
	Weekly	0	5.31	**39.4**
	Daily	0	1.22	49.5
Depression severity (if reported) (%)	Mild	0	**54.3**	12.1
	Moderate	0	37.9	**41.4**
	Severe	0	7.80	46.5

Table 17: Individuals in clusters (n) and proportion of population for "visual discrimination."

	Cluster 1	Cluster 2
n (%)	745 (84.9)	132 (15.1)

Table 18: Subjective clusters for "visual discrimination". Values correspond to the variable in first column. Excluding age, this is the proportion of individuals in the cluster who reported that characteristic.

NHANES Variable	Cluster 1	Cluster 2
Median Age	65	66
Reports difficulty seeing (%)	0	**100**

Table 19: Individuals in clusters (n) and proportion of population for "auditory discrimination."

	Cluster 1	Cluster 2	Cluster 3
n (%)	534 (60.8)	279 (31.8)	65 (7.4)

Table 20: Subjective clusters for "auditory discrimination". Values correspond to the variable in first column. Excluding age, this is the proportion of individuals in the cluster who reported that characteristic.

NHANES Variable	Level	Cluster 1	Cluster 2	Cluster 3
Median Age	-	62	68	76
Hearing quality without correction (%)	Excellent	**61.4**	0	0
	Good	38.6	0	0
	A little trouble	0	**59.1**	0
	Moderate trouble	0	40.9	0
	A lot of trouble	0	0	**90.8**
	Deaf	0	0	9.2
Wears a hearing device (%)	-	0.183	12.9	46.2

4.5 Discussion

In this section, the implications and limitations of this work are discussed.

4.5.1 Patient Characteristic Survey Outcomes

A survey was distributed to internal medicine physicians at the University of Maryland Medical Center to extract expert perceptions regarding the relationships between user characteristics and P&C case study task performance. For three high-level tasks, the physicians ranked characteristics based on the order they would consider them when evaluating expected patient task performance.

For tasks requiring physical effort, the characteristic that received the highest rating was “physical independence”. This seems logical, given that it summates a range of abilities that implicate independence in performing the expected task itself. Other highly rated characteristics include “disease severity”, “exercise”, “age”, “details of task”, and “disease history”. It is interesting that disease severity ranked higher than disease history. This is likely because well-controlled diseases, which are considered among a patient’s history, may not impact functionality, while those more clinically advanced diseases may contribute to debility.

For tasks requiring sensation and perception, “details of task” ranked highest. It is likely that internists envisioned a wide range of tasks that require sensation and perception while answering this question, so they responded that more specific characterization of the task would be required to predict success. Similarly, “details of task” was also highly rated for physical and cognitive tasks, however not as highly rated as for sensation and perception. Sensory modalities are fairly distinct, therefore it may have been more productive to ask about specific sensory pathways such as hearing or vision. Other highly rated tasks for sensation and perception were “disease history”, “age”, “attention”, “physical independence”, and “memory”. Disease history was likely rated highly due to the large range of impact disease may have on sensation and perception across various diseases. Similarly, prevalence of those conditions increases with age, which explains why internists would consider age as an important factor for these tasks. Surprisingly, physical independence was rated high for this task. While the nature of “a task requiring sensation and perception” may have been perceived as too vague, it is likely the experts consider “physical independence” to be influenced by sensory or perceptual deficits and had in mind diseases that affect physical and sensory systems together. Future iterations of this method will seek to refine the sensory task to elicit expert opinion more effectively.

For cognitive tasks, “memory”, “attention”, and “decision-making” were ranked the highest, as they are all integral elements of cognition. The next three highest ranked were “details of task”, “age”, and “psychiatric disorder”. Expert consensus was strongest for cognitive tasks, with the Borda count distribution appearing more skewed towards the top characteristics than for the physical and sensory/perception tasks.

There were some notable commonalities across tasks in the results. “Details of task” was rated highly for all three tasks included. This may indicate that experts would prefer more contextual detail when identifying important user characteristics. Other common characteristics to be included across tasks were “age” and “disease history” or “psychiatric disorder”, treating the latter two as equivalents. This points to the fundamental understanding of age being tied to human ability. The stated goal of this study was to include variables with direct links to task performance. While age itself does not change task performance, it is a strong predictor of other factors that may influence performance (Baker & Rogers, 2010; Fauth et al., 2017; Zimmer et al., 2015). Age is an easy variable to obtain and can provide substantial predictive power when data is unavailable for other performance-driving variables.

These results demonstrate a simple and efficient way to develop an understanding of domain-specific relationships between tasks and user characteristics. Medical professionals, in this case physicians, have extensive knowledge and experience that can be distilled into useful information with well-coordinated elicitation efforts. There are many challenges that make applying human factors principles difficult in medical device design (Saidi et al., 2019), but collaboration across disciplines could alleviate some of these challenges.

4.5.2 Patient Characteristics – NHANES Variable Mapping Results

For each task and each patient characteristics identified prior, variables were identified from NHANES to serve as metrics for the characteristic. For both physical tasks (gross upper-body movement, fine motor movement) NHANES contained many useful variables (Table 7; Appendix A, Table 56). Each characteristic was represented by at least one NHANES variable, and in most cases several more. Several variables were included that asked participants to rate their ability to perform certain tasks that represent specific cases of the generic task performance. Combined with more objective variables such as age and disease history, this data provides a rich picture of the capabilities of individual users.

Limitations were encountered with respect to sensory and perceptual characteristics. Only two NHANES variables were included for visual and auditory discrimination, age and reported ability (Appendix A, Tables 57-58). Because self-reported hearing quality was reported without correction from a hearing device, the variable "wears a hearing device" was included as well to verify that it was representative of typical hearing conditions. NHANES does contain data regarding audiometry, however these variables were additional self-reported hearing quality questions, which were judged to not provide enough additional information to justify incorporation. In past years, NHANES included auditory examination data, however it was determined that this would be difficult data to meaningfully interpret without specialized knowledge. Variables that NHANES did not include that would have been useful are the presence of common diseases for each sensory system (cataracts, glaucoma, etc.).

Cognitive user characteristics were also well represented in NHANES, with each characteristic being represented by at least one variable (Appendix A, Table 59). Psychiatric disorders were particularly well represented, including both the frequency and the severity of two disorders. Only one self-reported performance question was identified in “problems managing money”. While this task does encompass both low and high-level cognitive tasks, additional synonymous task performance questions would have been beneficial for further characterizing the self-assessed capabilities of the population.

4.5.3 Patient User Group Cluster Results

Data for each variable was acquired from the NHANES database for 2017-2018 and filtered to only include results for participants who reported diabetes. For physical and cognitive tasks, clusters were determined algorithmically. For sensory and perceptual tasks, clusters were determined subjectively.

Gross Upper-Body Movement: For “gross upper-body movement”, three dominant clusters were identified as shown in Tables 11 and 12. Common features across all three groups included age and exercise. In this work, individuals 65+ years are referred to as older adults, and those who are 18-65 as adults (Centers for Disease Control and Prevention, 2020b). The median age for all three clusters was older adult, and all three contained a majority of users who reported a sedentary lifestyle. The first and most predominant cluster primarily included users who do not report any physically inhibiting diseases and do not report difficulty performing tasks using the upper body. The second group reported some difficulty performing tasks with the upper-body. Further, arthritis, neck and back problems, and obesity were common for this group. Group three reported some difficulty performing multiple upper-body tasks and, like group two, commonly reported arthritis, neck and back problems, and obesity. Additionally, the majority of the individuals in this group reported three or more cardio-pulmonary conditions. It is suspected that group one would be the top task performer, group three would be the bottom, and group two would sit in-between.

Fine Motor Movement: For “fine motor movement”, two dominant clusters emerged as shown in Tables 13 and 14. When examining the composition of these groups, several similarities could be observed. Both groups consisted of older adults. For both groups, the only disease to be represented in the majority was arthritis. Both groups reported a sedentary lifestyle. In fact, when considering majority characteristics, the only real difference between these groups was that group two indicated some trouble grasping and holding small objects. This is a relatively important task for medical device usage, though, so it is expected that group two would exhibit poorer performance than group one.

Cognitive Tasks: For “applying existing knowledge” and “problem-solving and decision-making”, variables were identical, and results were presented together. Three clusters emerged for the cognitive tasks as shown in Tables 15 and 16. Groups were relatively diverse. The predominant trait across groups was reporting no difficulty managing money. Group one included primarily older adults and were not characterized by fatigue, confusion, memory, or decision-making. Further, this group did not exhibit psychiatric disorders. Group two was comprised of older adults and was not characterized by confusion, attention, memory, or decision-making. This group did, however, indicate some reoccurring fatigue. Additionally, the majority of this group reported anxiety and depression at least a few times a year. Severity of anxiety for most individuals was at least moderate, while for depression the severity was mild. For the final group, the median age was adult. The group reported significant fatigue, and problems with confusion, attention, memory, or decision-making. This group reported moderate, daily anxiety and reported at least moderate, weekly depression. It is suspected that group one would be the top task performer, group three would be the bottom performer, and group two would sit somewhere in-between.

Visual and Auditory Discrimination: Cluster variables for visual and auditory discrimination were only age and self-reported difficulty, therefore clusters were identified by grouping participants based on the self-reported variable. For visual discrimination, two clusters were created, with cluster one reporting no difficulty seeing and cluster two reporting difficulty (Tables 17 and 18). The median age for both groups was older adult. For auditory discrimination (Tables 19 and 20), the self-reporting variable consisted of 6-levels. Group one contained individuals who responded “excellent” or “good”. Group two stated that they had “a little trouble” or “moderate trouble”. Group three had either “a lot of trouble” or were deaf. Group one consisted primarily of adults, while group two and group three were primarily older adults. A minority of individuals for each group reported that they wear a hearing device, validating the assumption that self-reported quality of uncorrected hearing represented typical conditions for most group members.

4.5.4 Task-Specific Guidance

This section discusses specific guidance relevant to each task. For products requiring gross-upper body movements, efforts should be made to cater to the strength and mobility capabilities of dominant clusters. Specific gross upper-body movements may include lifting and holding components for repeated and extended periods of time. Products should be designed to minimize the time required for these movements, and the number of times they must be repeated (Marras, 2012).

Most patient-facing medical devices will require some fine motor movement, for example pressing buttons, touching a screen, twisting or pinching components, etc. Efforts should be made to ease device use for users who have difficulties with these types of movements. For example, for devices with screens that require scrolling, key content should condensed as much as possible to reduce the amount of scrolling required (L. C. Li et al., 2013). As another example, unique and repetitive finger manipulations should be minimized. Individuals with arthritis and those with limited hand dexterity have indicated that minimum or no-button device designs are preferable (Domańska et al., 2017).

Perception of visual and auditory information is also very important in medical device interaction. A common strategy to accommodate deficiency in either category is to map the device output modality to the opposite or a different output modality (Dascalu et al., 2017; Niazi et al., 2016). For visual discrimination, the at-risk group identified encompasses a broad swath of visual capabilities. The differentiating variable only identified subjects who "had difficulty seeing". This includes subjects with mild visual disability to subjects with blindness. The strategy to accommodate this range may vary (Siebra et al., 2015). To ensure complete coverage of user needs, the need for visual discrimination can be eliminated entirely, for example with the inclusion of text-to-speech features (Tomlinson, 2016). A more conservative approach, that might risk disaccommodating certain portions of the group, could include ensuring all screens and text are large and highly contrasted. The device itself should be large and locations where the device are held should be made obvious (Heinemann et al., 2016).

For auditory discrimination, the trade-offs to accommodate different types of deficiencies can be more precisely examined. Because there were two groups with two levels of hearing impairments (mild-moderate, severe-deaf) identified, the consequence of not including certain features can be quantified. For the mild-moderate group, it may only be necessary to tailor auditory output to their specific needs. This could include increasing the minimum device volume or minimizing voice output. Non-speech sounds may be preferrable because they convey information quicker and simpler than speech sounds (Shoaib et al., 2020). For deaf individuals, auditory output should be mapped entirely to other modalities, primarily visual or tactile. Captions and sign language transcriptions can be leveraged when possible (Siebra et al., 2017). This type of intervention becomes more complicated when there is impairment of multiple sensory channels, however this was not considered in this work.

Numerous cognitive tasks are required for medical device usage. Cognitive tasks can range from recalling the meaning of an alert, to following procedures for device use, to evaluating and acting on diagnostic output. For the groups identified in this analysis, different interfaces could be offered. For groups 2 and 3, device output should be simplified, and aid provided for decisions with potential health-related consequences. It may be beneficial, however, to provide the highest performing group greater perceived control over the output of the device, as this has been shown to increase willingness to use medical devices in some cases (Princi & Krämer, 2020). Further, cluttered and complex interfaces can create difficulties for certain users, particularly those identified in risk group 3 (Wildenbos et al., 2018). Device operation procedures and interface navigation should be made as simple as possible without inhibiting device functionality (Roman et al., 2017). Patients with compromised cognition may also have difficulty with tasks whose performance is traditionally linked to health literacy levels. Efforts should be made to minimize the demand on literacy by providing clearly stated content with accessible language (Czaja et al., 2015).

Across all tasks, features to accommodate high-risk user groups can be informed by existing guidelines. Recent clinical guidelines have suggested the inclusion of several features across diabetes self-management technology (Chomutare et al., 2011). Examples include education and personalized feedback, weight management, psychosocial care, and medication management. Many of these guidelines apply beyond the diabetes context to other medical devices as well. Designers should consider how these features can be incorporated into patient-facing medical devices, and how the features can be personalized or prioritized given the characteristics of the user group. For example, psychosocial care features could be prioritized and expanded for individuals in groups at high-risk for cognitive errors.

4.5.5 Proposed Use

The purpose of identifying task-specific user groups was to provide a basis for targeted human factors evaluations and subsequent design personalization where optimizing human performance is the goal. Identifying groups based on characteristics linked to task performance can aid in ensuring prominent, homogeneously performing users are represented in the design process. The groups identified should represent clusters of individuals who are expected to perform tasks similarly. Utilizing existing data to define user groups is more efficient than surveying or engaging with new participants to model performance groups. This makes this approach especially useful for medical device manufacturers, who are typically operating with constrained resources.

After performing the presented methodology, the next step should be to quantify task performance for each task-specific user group generated. Task performance can then be used to justify inclusion/exclusion of features for product variants. In Chapter 5 of this dissertation, performance for P&C Physician Judgment

Tasks will be quantified for each user group utilizing expert elicitation. In Chapter 6, user group task performance quantities will drive product concept differentiation.

While expert elicitation is the suggested means to quantify performance in this dissertation, there are other means outside of this scope. If empirical testing is possible, formative and summative usability analyses should be performed. Clustered groups can serve to guide targeted recruitment of dominant user strata. Recruitment stratification goals can be set using the cluster proportions identified. By monitoring the progress of these goals, recruitment strategies can be adjusted during a study to ensure representation. These groups could also be used post-study to evaluate the adequacy of recruitment and determine if a study extension is required.

Output user groups could also be used outside of the design context. Risk analysts, systems engineers, and technical specialists in the medical context could use this approach to develop models for use-cases and to model certain risk scenarios. Health advocates and public health professionals could use this approach to develop targets for public health initiatives.

4.5.6 Limitations

This work had several limitations. The proposed approach requires access to domain experts for input to identify relevant tasks and user characteristics, otherwise the researchers must have the domain expertise themselves. There were several limitations with respect to how sensory and perceptual tasks were framed in the user characteristic survey that potentially could have been avoided with more care. Additionally, utilizing existing data limits the user characteristics that can be included in the clustering. For some applications, there may be critical characteristics that are not available in a database like NHANES. In this paper, NHANES had limited data for visual and auditory task characteristics. A further limitation is that the characteristics available for clustering will not be able to explain all likely variance in task performance. The actual range of performance that would be observed for a given user group would likely be wide, though hopefully with the expert-driven steps taken prior, performance will be clustered as homogenously as possible. Finally, this approach does not account for correlation between patient group membership. It is possible that there may be patterns to patient group membership across tasks.

Chapter 5: Quantifying Human Performance for Heterogeneous Population User Groups using Expert Elicitation

5.1 Introduction

In this chapter, an expert-driven approach for quantifying human performance (risk of user error) in heterogeneous user populations is proposed. The approach uses input from Chapter 4 (P&C Physician Judgment Tasks and task-specific user groups) and outputs probability distributions for task success for each task-specific user group (Figure 4). The purpose of this approach is to supplement traditional in-person human factors testing with heterogeneous user populations to mitigate some of associated resource burden. The output will also serve as input to the optimization model proposed in Chapter 6. The approach is first introduced in a generalized form that is intended to be applicable to any domain. Then, the method is demonstrated on the diabetes patient self-management device case study.

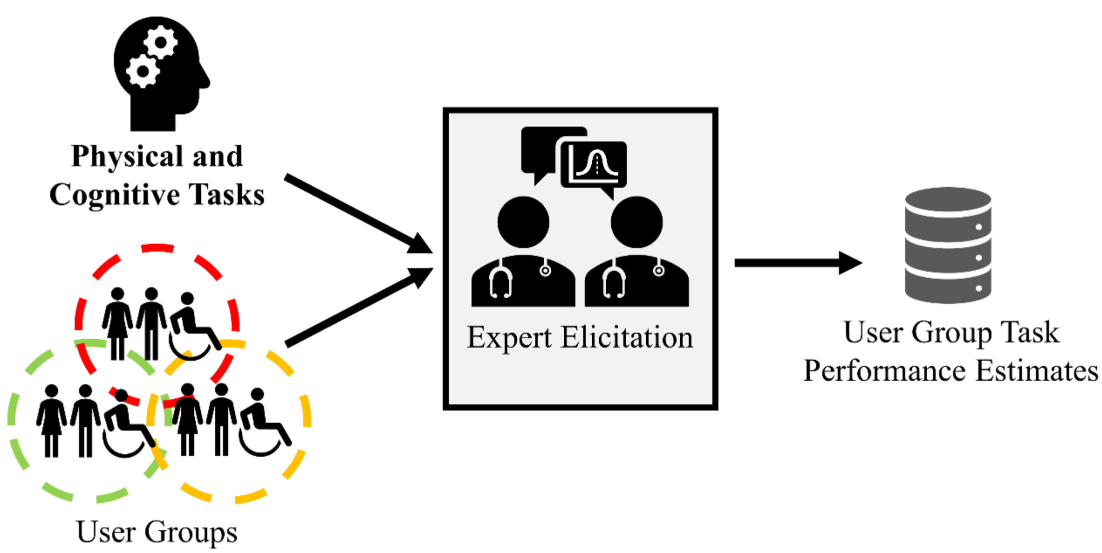


Figure 4: Summary of Chapter 5.

5.2 Methodology

The approach to be used in this chapter is expert elicitation, where desired variables are estimated by experts as probability distributions and aggregated as a combined "decision-maker" (Colson & Cooke, 2018). In this approach, internal medicine physicians are targeted as domain experts for participation (see Chapter 4.1 for justification).

5.2.1 Elicitation Protocol

To conduct the expert elicitation, there are several decisions that must be made regarding the format of the elicitation and the design of questions. For this methodology, the Cooke protocol is used (Cooke, 1991). It is critical that physician participants can provide input at their convenience because medical providers have very busy schedules. If not, participation may suffer, especially if no direct incentives are offered (as was the case in this work). The Cooke protocol is ideal because it does not require participants to convene and can be completed at any time.

In the Cooke protocol, prior to estimating quantities of interest, experts are asked to make judgments on domain-relevant "calibration questions" with known values. Experts are evaluated on these calibration questions, and their performance is used to weight and aggregate responses for the quantities of interest. For this methodology, judgments for all questions are elicited as a 5th percentile, a 50th percentile, and a 95th percentile. This provides a distribution that the expert is 90% confident the true value is contained.

Calibration questions are questions with known values, in the same format as the elicitation questions (Colson & Cooke, 2017). Elicitation questions and calibration questions should be thematically linked with the domain, such that all target experts are roughly equally qualified to answer them. These questions should not test the expert's domain-expertise, rather they should test how well an expert can express a judgment translated to a probability distribution (Dias et al., 2018). If the researcher has prior data, these questions could be derived from this data. These questions could also be domain relevant statistics obtained from reports or literature.

5.2.1.1 Evaluating Experts

Calibration questions can be evaluated using two different metrics – calibration and information. Calibration measures the statistical accuracy of the judgments, and information measures the confidence of the judgments (Colson & Cooke, 2018). A high performing expert will have high accuracy and high confidence (i.e., smaller confidence intervals).

Calibration Score: Calibration score compares the distribution of true values within the intervals provided with the expected distribution of values. For a perfectly calibrated expert, it is expected that 5% of true values will fall below the 5^{th} percentiles of the provided intervals, 45% will fall between the 5^{th} and 50^{th} percentiles, 45% will fall between the 50^{th} and 95^{th} percentiles, and 5% will fall above the 95^{th} percentiles. For example, if 20 calibration questions were given, perfect calibration would be achieved if the true values were distributed as [1,9,9,1], following the order of intervals described.

Given this, calibration can be measured using Kullback-Leibler (KL) divergence (aka relative entropy), a metric grounded in information theory and used to measure the difference between a probability distribution and a reference distribution. In this case, the reference probability distribution is p = [0.05,0.45,0.45,0.05]. KL divergence is formatted as follows in Equation 1 (Dias et al., 2018):

$$I\,(s,p) = \sum_{i=1}^{n} s_i \ln\left(\frac{s_i}{p_i}\right) \tag{1}$$

where s_i is the observed proportion of values in interval i, p_i is the expected reference proportion of values in interval i, and n is the number of intervals.

In Cooke's protocol, an expert is scored given a statistical hypothesis, where the null hypothesis is that the inter-quartile intervals containing the true values for calibration questions is drawn from the reference probability distribution. The divergence metric can be formulated such that it is approximated as the chi-square distribution as in Equation 2 (Dias et al., 2018), therefore allowing a p-value to be obtained from a chi-square goodness-of-fit test.

$$Calibration\ = \Pr\{2qI(s,p) \leq x\} \rightarrow\ 1 - \chi^2_{n-1}(x)\,, as\ q \rightarrow\ \infty \tag{2}$$

where q is the number of calibration questions given and χ^2_{n-1} is the cumulative distribution function of the chi-square distribution. Calibration score can then be interpreted as the probability of seeing a more extreme divergence metric given the provided and reference distributions. Scores range between 0 to 1, with 1 being the best possible score.

Information Score: Information score compares the provided confidence intervals to the background range of values intrinsic to the question. Information score is necessary because, hypothetically, an expert could achieve perfect calibration without providing a useful or informative judgment by specifying large confidence intervals. In typical practice, the intrinsic range for each questions is determined by taking the min and max values on either end of the range across experts and adding a small "overshoot" equal to 5-10% of the min and max values (Dias et al., 2018). 5% was used in this work. An informative response will encompass a relatively small proportion of the intrinsic range. Again, KL divergence is leveraged to compare the distribution of the inter-quartile intervals provided against a uniform distribution across the entire intrinsic range. When 3 percentiles are elicited (i.e. 4 intervals), it can be calculated as shown in Equation 3 (Dias et al., 2018).

$$Information\ score\ = 0.05 \ln\left(\frac{0.05}{x_{i1} - x_{i0}/x_{i4} - x_{i0}}\right) + 0.45 \ln\left(\frac{0.45}{x_{i2} - x_{i1}/x_{i4} - x_{i0}}\right) + \ 0.45 \ln\left(\frac{0.45}{x_{i3} - x_{i2}/x_{i4} - x_{i0}}\right) + 0.05 \ln\left(\frac{0.05}{x_{i4} - x_{i3}/x_{i4} - x_{i0}}\right) \tag{3}$$

where x_{i0} and x_{i4} are the lower and upper bounds for the intrinsic range of calibration question *i*. x_{i1}, x_{i2}, *and* x_{i3} are the expert provided 5th, 50th, and 95th percentile for calibration question *i*.

Given *information score* and *calibration score*, a final score for each expert is calculated as shown in Equation 4. An expert with a higher score is determined to be a "better" expert for making domain relevant judgments.

$$Score = \frac{\sum_{i=1}^{q} Information_i}{q} * Calibration\ score \tag{4}$$

It is worth noting that unlike calibration score, information score can be evaluated without knowing the true value of the question. Therefore, expert confidence can be examined using information score for the elicitation questions that do not have known values.

5.2.1.2 Combining Expert Judgments

To utilize the elicited values practically, a method for combining the estimates into a single decision-maker is required. Theoretically, this combination should provide a best guess estimate for the unknown quantities. In practice, there are typically three ways to determine the combined decision-maker, as summarized in Table 21 (Dias et al., 2018). In this work, all three combinations are presented and compared.

Table 21: Methods for combining estimates into a final decision-maker.

Equal Weighted	Elicited values are aggregated across all experts.
Performance Weighted	A weighted aggregation is performed for all elicited values, where the weight for each expert is their score (Eq. 4) normalized across all experts.
Optimized Weighted	Weighted aggregations are calculated where experts are eliminated from the weighted aggregation if their calibration score falls below a cutoff value α. Starting at the lowest calibration score in the analysis, α is incremented and a weighted aggregation is calculated each time an expert is removed. Each combined aggregation is re-tested on the calibration questions, and the combination of experts with the highest score (eq. 4) is determined to be optimized.

While equal weighted considers each expert as equivalent, performance weighted gives more weight to predictions of more reliable experts (higher calibration score). The optimized weighted approach is an extension of performance weighted that removes the least calibrated expert one-by-one in an iterative manner to determine the optimal combination of experts that maximizes expected performance. While Cooke's protocol suggests aggregating estimates by weighting and combining fitted cumulative distribution functions is the best option, aggregation in this work is performed by simply averaging elicited quantiles. The methodology in this work is intended for engineers to reproduce without specialized statistical knowledge and experience. Aggregation of estimates by averaging quantiles is a computationally simple approach that can be reproduced by engineers without prior experience performing expert elicitation (Lichtendahl et al., 2013).

5.2.2 Generating Elicitation Questions

To define the variables to be quantified in the expert elicitation, three elements should be defined: **1) Product Tasks**; **2) Contextual Task Details**; and **3) Task-Specific User Groups**.

Product Tasks: For the proposed method, it is suggested that generalized physical and cognitive tasks are defined for the system. Modeling tasks as general allows results to be used in future design validations and new system designs for the same population of users. It also facilitates continued model validation. Tasks that are too specific may limit their generalizability and therefore future designs may require additional elicitation efforts, thus requiring additional resources. **In this methodology, P&C Physician Judgment Tasks identified in Section 4.3.1 are used.**

Contextual Tasks Details: While keeping tasks general does promote reuse, it does limit how precisely they can be applied to specific situations. As such, it may be necessary to create sub-variables for tasks given contextual information. For example, the performance of auditory discrimination may depend on the entity being discriminated and may be too broad for experts to make precise estimates. Instead, experts could be asked to estimate performance for discriminating speech as well as non-speech sounds. It is important to consider what contexts may be encountered for the specific application. For example, in certain systems, speech sounds may be irrelevant, and this variable could be eliminated.

Task-Specific User Groups: For experts to make estimates about task performance, they will require information about the population of individuals performing the task. For heterogeneous populations, these characteristics are highly varied and, therefore, users must be grouped and their task-relevant user characteristics specified for each task prediction. This process of identifying task-relevant user characteristics will produce more homogenous groups in terms of task performance, and thus experts can make more confident estimations (Privitera, 2020). This also ensures that performance heterogeneity is captured for all tasks, and that a single task associated with particularly heterogeneous user characteristics does not overshadow other, less heterogeneous tasks. **In this work, the task-specific user groups identified in Section 4.4.3 are used.**

Practical Considerations: As the elicitation question elements (P&C Physician Judgment Tasks, Contextual Task Details, Task-Specific User Groups) and calibration questions are being defined, it is important to consider how they will affect the elicitation length. If resources are limited or incentives for participants are not available, it is critical to minimize survey length to ensure adequate participation. The number of values that must be elicited can be calculated as shown in Equation 5.

$$n = a\left(\sum_{i=1}^{T} c_i m_i + q\right) \tag{5}$$

where a is the number percentiles elicited for each variable, T is total number of tasks for the elicitation, c is the number of contexts for task i, m is the number of user groups for task i, and q is the number of calibration questions.

Question Design: The format of the questions and design of the question dissemination medium are key considerations for question design. The format of the questions and how they are presented should conform with the expert's mental model of the population and the tasks being investigated. A formal or informal requirements elicitation prior to eliciting judgments can facilitate development of the elicitation material. There are several questions that should be answered during this process. When eliciting risk estimates, the quantity can be elicited as either a probability or a proportion. For example, a question could be worded as "what is the probability an individual will fail to perform task x" or "out of [10, 100, 1000] individuals, how many would you expect to fail to perform task x". Further, the expert pool may have linguistic preferences regarding the names of tasks. The terminology used to describe a task may connote unintentional meaning in certain social or professional contexts.

5.3 Case Study Application

This section describes the application of the proposed approach to the diabetes population case study. The expert elicitation method was used to quantify risk for diabetes patients for several generic tasks that are integral to medical device interaction. The demonstration did not focus on a specific product or group of devices, instead providing generalized predictions for interactive device use by diabetes patients. However, the most relevant and useful application of these predictions are in medical device use.

Experts were recruited from the University of Maryland Medical Center (UMMC) located in Baltimore, Maryland. Eligibility criteria included UMMC employment as an internal medicine physician. 6-12 experts were sought as guidance suggests that this will provide study robustness while avoiding diminishing returns (Knol et al., 2010). IRB approval was sought, and the study was exempted from IRB review (IRB Package # 1559401-1).

5.3.1 Elicitation Question Generation

To determine the format of the quantities to be elicited, an informal requirements elicitation was conducted with a sample of physicians. Several important preferences were discovered. First, experts indicated that they consider patient performance as the likelihood of success. Therefore, questions were formatted to ask for the probability of task success instead of failure. Next, experts were polled on their preferred response format. The following options were provided:

a. A patient in population X performs some task. What is the probability they will successfully perform this task?
b. 100 patients who belong to population X perform a task. How many patients will successfully perform this task?
c. 10 patients who belong to population X perform a task. How many patients will successfully perform this task?

Results of the poll indicated that response format b was the preferred format and was selected for the elicitation. If this approach is applied to a different user population, this same poll can be used, where “patient” is replaced to reflect the use case (e.g., “operator”).

5.3.1.1 Calibration Questions

Five calibration questions were generated for the elicitation. Questions were derived from statistics in the Centers for Disease Control (CDC) 2020 National Diabetes Statistics Report (Centers for Disease Control and Prevention, 2020a).

Percentages were rounded to whole numbers to reflect the “Out of 100 patients” question format. Questions and their values are shown in Table 22.

Table 22: Calibration questions, abbreviations, and correct values.

#	Calibration Question	Abbreviation	Value
1	Out of 100 patients, how many US adults with diagnosed diabetes would you expect to have a BMI greater than 25 kg/m^2?	BMI	89
2	Out of 100 patients, how many US adults with diagnosed diabetes would you expect to do at least 150 minutes of physical activity per week?	Exercise	16
3	Out of 100 patients, how many US adults with diagnosed diabetes would you expect to have a non-HDL level of 130 mg/dL or higher?	NonHDL	44
4	Out of 100 patients, how many US adults with diagnosed diabetes would you expect to have an A1C value of greater than 9.0%?	A1C	15
5	Out of 100 patients, how many US adults with diagnosed diabetes would you expect to also have chronic kidney disease (at any stage)?	Kidney Disease	37

5.3.1.2 Task Performance Questions

As mentioned prior, there were three components required for each question: the task, details of the task, and task-specific user groups. This section details the generation of each for the demonstration.

Tasks used were those generated in Chapter 4. Tasks are listed in column 1 of Table 23. Lack of contextual information for each task creates additional uncertainty regarding how it will be performed. For each task, contextual details were specified such that experts would be able to make more confident estimations. Details were specifically tailored for the medical device use context. For each question, specific examples were identified to be presented to participants to demonstrate the context. Task details for each task are summarized in Table 23.

Table 23: Tasks, task details, and examples presented to participants.

P&C Case Study Task	Task-Type	Detail of Task	Examples Provided to Experts
Gross Upper-body Movement	Physical	Requires low exertion	• Lifting or carrying a small object or device • Raising the arms above the head • Pulling a door open
Fine Motor Movement		Requires low exertion	• Pressing a button • Twisting a component into place • Grasping a device
Visual Discrimination	Sensory	Simple	• Detecting a flashing light • Discriminating between colors
		Complex	• Identifying details on a phone screen • Reading small print
Auditory Discrimination		Non-speech	• Detecting beeps or alarms from a device
		Speech	• Discriminating speech from a device
Applying Existing Knowledge	Cognitive	Simple	• Recalling instructions • Recalling values
		Complex	• Performing a set of procedures • Enacting instructions for device operation
Problem-solving and Decision-making		Simple	• Determining if a diagnostic reading is within a pre-defined normal range • Determining if food meets pre-defined dietary guidelines
		Complex	• Determining a course of action given a new device warning message • Determining a course of action given a new health symptom

Task-specific user groups were taken from Chapter 4. For each patient user group, statistics were translated into plain-word descriptions to be presented during the elicitation. Groups and descriptions are displayed in Table 24. Group number corresponds to increasing subjective patient risk.

Table 24: Task-specific patient user groups and their plain-word description.

Task	Group #	Dominant User Group Characteristics
Gross Upper-body Movement	1	Older adult; Physically independent; Sedentary lifestyle
	2	Older adult; Partially physically dependent; Sedentary lifestyle; Arthritis; Neck and back problems; Obesity
	3	Older adult; Partially physically dependent; Sedentary lifestyle; Arthritis; Neck and back problems; Obesity; At least 3 cardiopulmonary conditions
Fine Motor Movement	1	Older adult; Physically independent; Arthritis; Sedentary lifestyle
	2	Older adult; Partially physically dependent; Arthritis; Sedentary lifestyle
Visual Discrimination	1	Older adult; With best-corrected vision, does not report difficulty seeing
	2	Older adult; With best-corrected vision, reports some difficulty seeing
Auditory Discrimination	1	Adult; Does not report difficulty hearing; Does not use hearing aid or other listening devices
	2	Older adult; Reports some/moderate difficulty hearing; Does not use hearing aid or other listening devices
	3	Older adult; Deaf or reports significant difficulty hearing; Does not use hearing aid or other listening devices
Applying Existing Knowledge	1	Older adult; Reports normal energy levels; Reports normal attention, memory, and decision-making
	2	Older adult; Reports low energy sometimes; Reports normal attention, memory, and decision-making; Experiences moderate anxiety and/or depression symptoms intermittently
	3	Adult; Reports low energy constantly; Reports impaired attention, memory, and decision-making; Experiences severe anxiety and/or depression symptoms constantly
Problem-solving and Decision-making	1	Older adult; Reports normal energy levels; Reports normal attention, memory, and decision-making
	2	Older adult; Reports low energy sometimes; Reports normal attention, memory, and decision-making; Experiences moderate anxiety and/or depression symptoms intermittently
	3	Adult; Reports low energy constantly; Reports impaired attention, memory, and decision-making; Experiences severe anxiety and/or depression symptoms constantly

To summarize, when all three elements are combined, the result is 1-2 task details per task, and 2-3 task-specific user groups performing each task. This resulted in 27 task estimations plus five calibration questions for each physician participant.

5.3.1.3 Background Questions

Finally, in addition to the task estimations and calibration questions, several background questions were developed for the experts, with the intent of modeling expert performance with professional background to facilitate future recruiting efforts. In addition, the influence of expert bias (experience, patient risk level) was assessed using this information.

These questions included:

1. How many years of experience do you have as a practicing physician?
2. Considering your entire career, please estimate the following regarding the patients you have provided care for:
    - % of patients with Medicare or Medicaid
    - % of patients who were uninsured
3. Considering your entire career, have you spent more time providing inpatient or outpatient care?

5.3.2 Elicitation Survey Development

Given the question components identified prior, survey questions were developed. The questions were implemented in Qualtrics (Qualtrics XM, 2020). Several rounds of testing and refinement were conducted with the survey by the researchers. Questions were presented to participants as sliders, constrained to whole numbers from 0 to 100. For each question, responses were requested ordered as 5th percentile, 50th percentile, and 95th percentile. The following instructions were provided for interpreting the requested values:

"For each quantity, you will be asked to estimate:

1. 5th Percentile - a value such that there is only a 5% chance the true value is smaller.
2. The Expected Quantity - the value with the highest probability of being true.
3. 95th Percentile - a value such that there is only a 5% chance the true value is larger.

The 5th and 95th percentile will define a range that you are 90% confident the true value is contained. These values should not overlap with the expected value."

For each task and task detail, examples from Table 23 were presented. For each question, the details for the patient user group were presented. Each question was presented with the following format: "100 patients from Patient Group ## perform a **[TASK]**. The task requires **[DETAIL OF TASK].** How many will succeed at this task?" An example question presentation is shown in Figure 5.

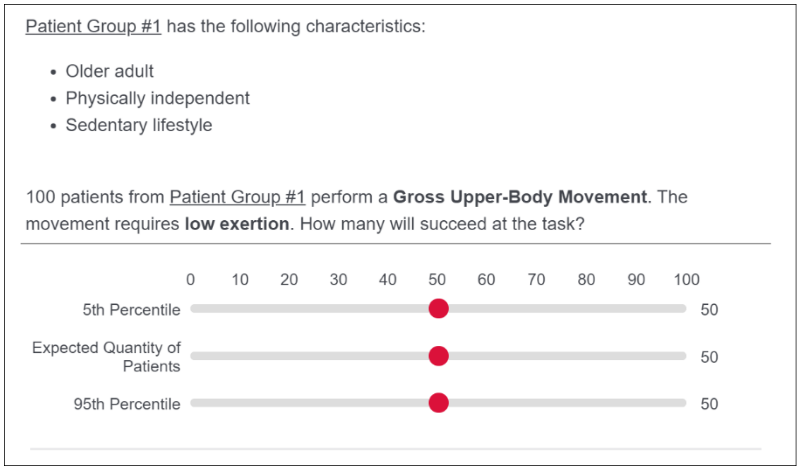


Figure 5: Example question presentation from elicitation survey.

For each task, questions were ordered based on increasing subjective user group risk-level as evaluated by the researchers (corresponding to the order in Table 24). In this context, risk-level refers to increasing likelihood of failure for the corresponding task. While randomization is commonly used to remove ordered bias in surveys, it was not done in this work. Rather, questions were ordered to support physician decision-making regarding patient risk. Evidence shows that, when estimating relative magnitudes in the medical context, presenting the information in an ordered manner can make judgments regarding those magnitudes easier (Reyna, 2008).

Restrictions were implemented to prevent participants from specifying any of the values as equal, and to require the 5th percentile < expected value < 95th percentile. Further, participants were not allowed to advance through the survey without adjusting each slider at least once.

5.3.3 Analysis

Several statistical analyses were performed to explain the results. Linear mixed-models with random intercepts were fit to examine relationships between background questions and performance outcomes. Background questions were included as independent variables and were modeled with elicited expected values for all task predictions, information score for all task predictions, and the combined scores for all calibration questions. For all analyses, participant ID and task-task detail pairs were treated as random effects.

Further, expected values, ignoring the elicited 5^{th} and 95^{th} percentiles, and information scores were compared for various groupings of the task estimates to better understand how experts approached the estimation task. Expected values and information scores were compared across risk groups, where groups were coded as 1-3 based on their order shown in Table 24. For tasks with only two associated groups, group 2 was coded as risk-level 3. For these analyses, risk-level was the independent variable and expected values and information score were dependent variables. Participant ID and task-task detail pairs were treated as random effects. Following this, post hoc multiple comparisons was performed using Tukey's method.

Finally, to identify fundamental differences in decision-making regarding type of task, expected values and information scores were examined with tasks grouped at their highest-level – physical, sensory, and cognitive. Again, participant ID was treated as a random effect. Post hoc multiple comparisons were performed using Tukey's method.

5.4 Results

The following section contains the following results: 1) Summary Statistics; 2) Calibration Question Results; 3) Task Estimation Results; 4) Risk-Level Analysis; and 5) Task-Type Analysis.

5.4.1 Summary Statistics

Twelve internal medicine physicians voluntarily completed the elicitation questionnaire. Table 25 summarizes the background question responses for all participants. Experts had a wide range of years of experience, with most serving high-risk patients in an inpatient setting.

Table 25: Summary of background question responses for experts.

Expert #	Years of Experience	Est. % Patients with Medicaid or Medicare	Est. % Patients Uninsured	Majority Inpatient vs. Outpatient
1	3	30	15	Outpatient
2	14	60	20	Inpatient
3	38	50	19	Outpatient
4	15	50	10	Outpatient
5	9	65	30	Inpatient
6	16	85	20	Inpatient
7	11	40	10	Outpatient
8	1	61	41	Inpatient
9	4	71	20	Inpatient
10	15	66	47	Inpatient
11	6	51	10	Inpatient
12	3	66	17	Inpatient
Mean (SD)	11.25 (9.58)	57.92 (14.1)	21.58 (11.5)	-

5.4.2 Calibration Question Performance

Figure 6 displays expert responses to calibration questions, with the correct value indicated with a vertical blue line. Full question text is shown in Table 22.

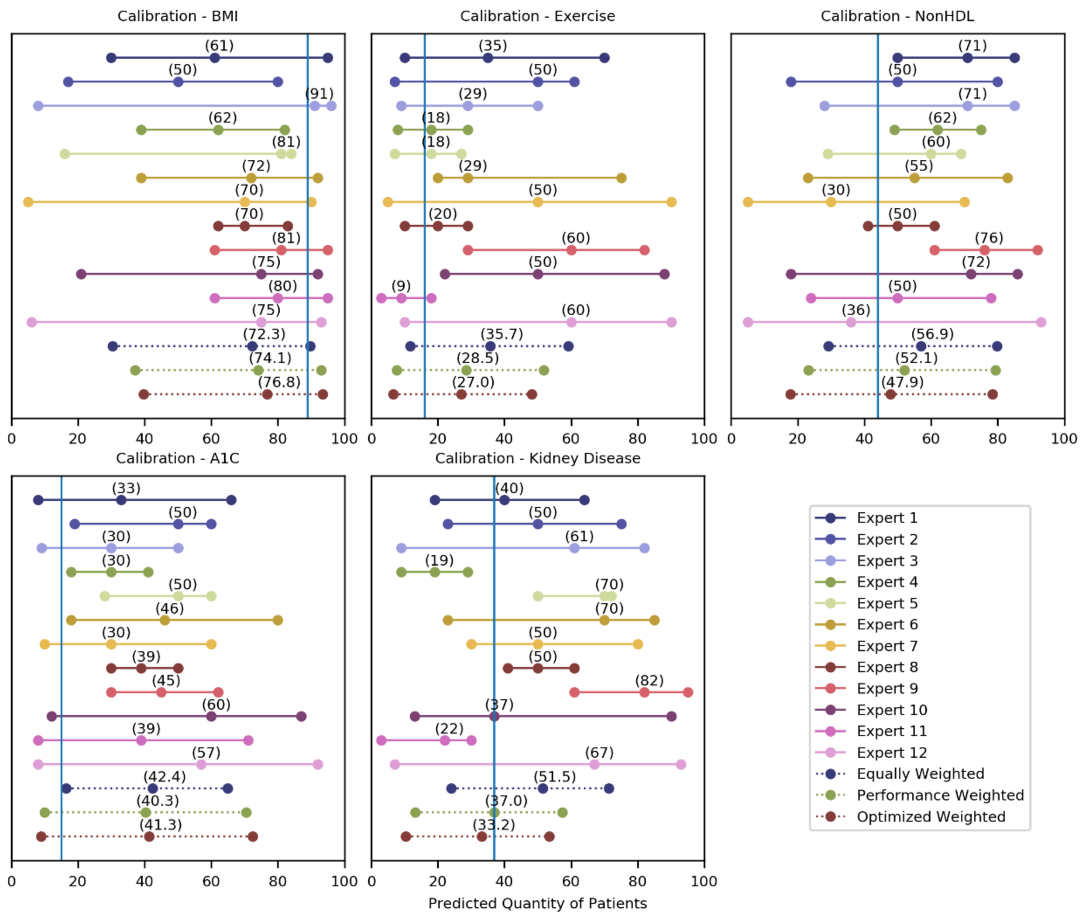


Figure 6: Elicitation results for calibration questions. Correct value indicated by vertical blue line. Full question text is in Table 22.

Table 26 displays expert performance for the calibration questions. Weight is the normalized score across all experts to be used in the "Performance Weighted" combinations described in Table 21. Optimized weight is the weight determined following the optimization procedure discussed in Table 21. A zero in this column indicates this expert was eliminated. The final alpha value used was 0.608. Experts 7, 10, 11, and 12 were the highest performing experts.

Table 26: Expert calibration question performance.

Expert #	Mean Information Score	Calibration Score	Score	Weight	Optimized Weight
1	0.357	0.411	0.147	0.141	0
2	0.359	0.064	0.023	0.022	0
3	0.458	0.046	0.021	0.020	0
4	0.932	0.002	0.002	0.002	0
5	1.015	0.014	0.014	0.014	0
6	0.318	0.101	0.032	0.031	0
7	0.223	0.740	0.165	0.159	0.211
8	1.136	0.014	0.016	0.015	0
9	0.662	0.0001	0.0001	0.0001	0
10	0.180	0.608	0.110	0.106	0.140
11	0.745	0.608	0.453	0.436	0.578
12	0.076	0.740	0.056	0.054	0.072

Table 27 contains the calibration question performance for the equally weighted, performance weighted, and optimized weighted "decision-maker." The optimized weighted decision-maker achieved the highest score, with performance weighted close behind.

Table 27: Calibration question performance for combined decision-makers.

Decision-maker	Mean Information Score	Calibration Score	Score
Equally Weighted	0.363	0.411	0.149
Performance Weighted	0.348	0.740	0.257
Optimized Weighted	0.358	0.740	0.265

Table 28 summarizes the performance on each calibration question. "Frequency correct" refers to the number of times the correct answer fell within the elicited confidence intervals.

Table 28: Average performance by calibration question.

Question Abbreviation	Frequency Correct	Mean Information (SD)
BMI	8	0.528 (0.336)
Exercise	9	0.514 (0.502)
NonHDL	9	0.593 (0.340)
A1C	6	0.475 (0.345)
Kidney Disease	7	0.423 (0.492)

Full question text shown in Table 22.

Question performance was similar, with “How many US adults with diagnosed diabetes would you expect to have an A1C value of greater than 9.0%” falling within the specified intervals the least. There were no significant differences between the specified information per question provided.

5.4.3 Task Elicitation Results

The results of the task performance elicitations are shown below. Figure 7 displays results for gross upper-body movement and fine motor movement. When looking at combined values, quantites for task success followed a logical pattern for both tasks, decreasing with subjective risk-level. Gross upper-body movements were percieved as a higher risk task.

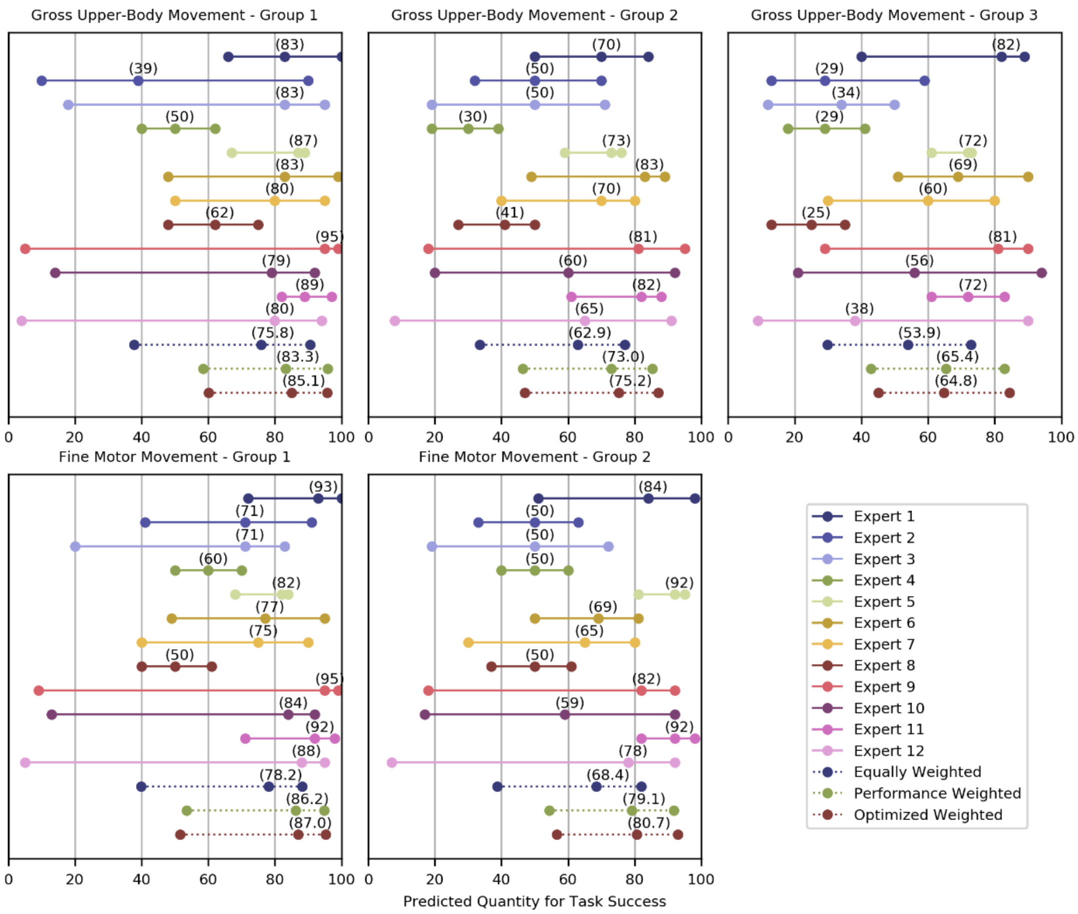


Figure 7: Elicitation results for gross upper-body movement and fine motor movement tasks.

Figure 8 displays results for visual discrimination. As expected, task success decreased with increasing user group risk. Experts also specified increasing risk with increasing task complexity.

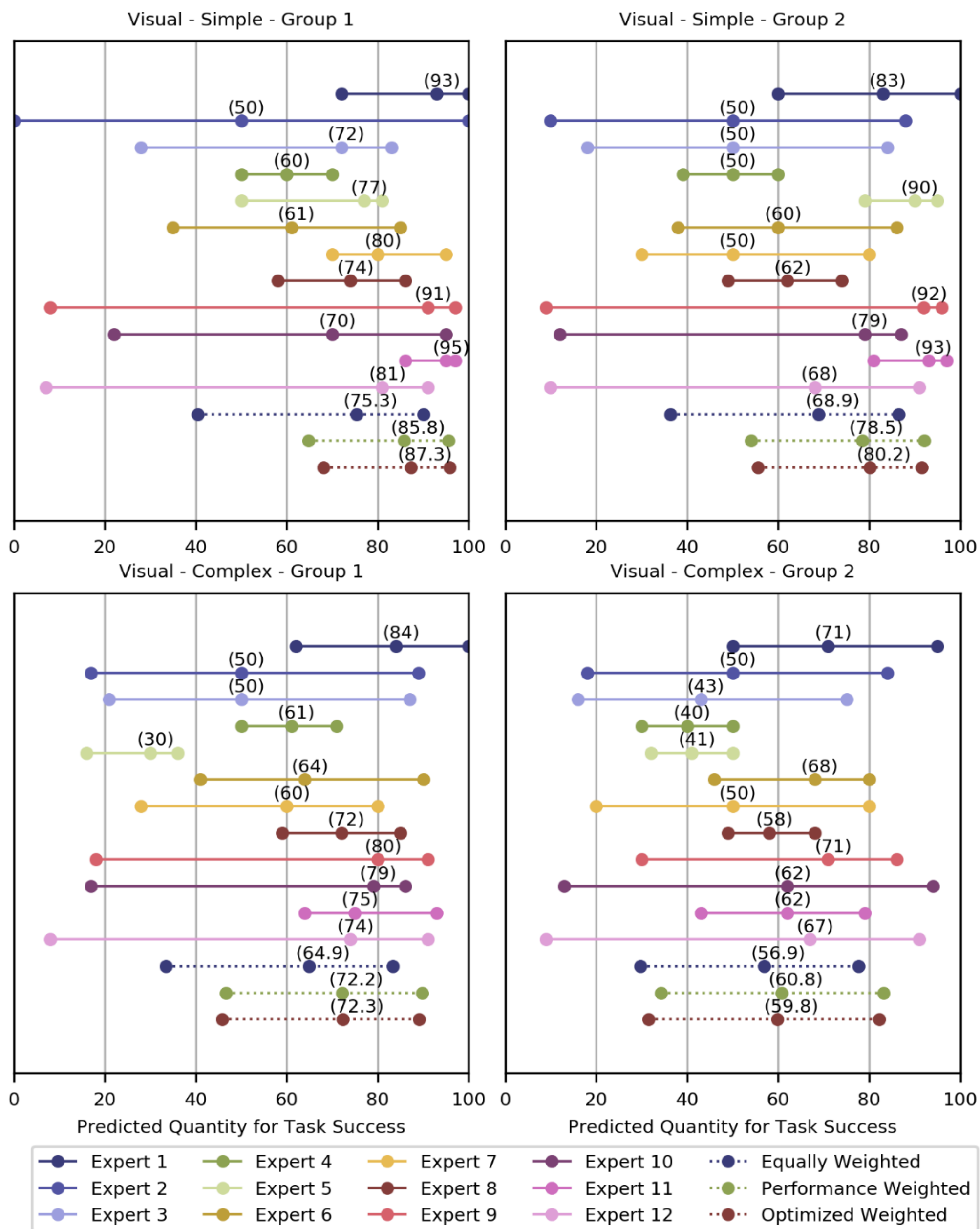


Figure 8: Elicitation results for visual tasks.

Figure 9 contains results for auditory task performance. As before, the combined decision-makers followed a logical pattern, with increasing risked being elicited with increasing risk-level and task complexity. Detecting and perceiving speech sounds for group 3 was perceived to be the highest risk task-user group combination.

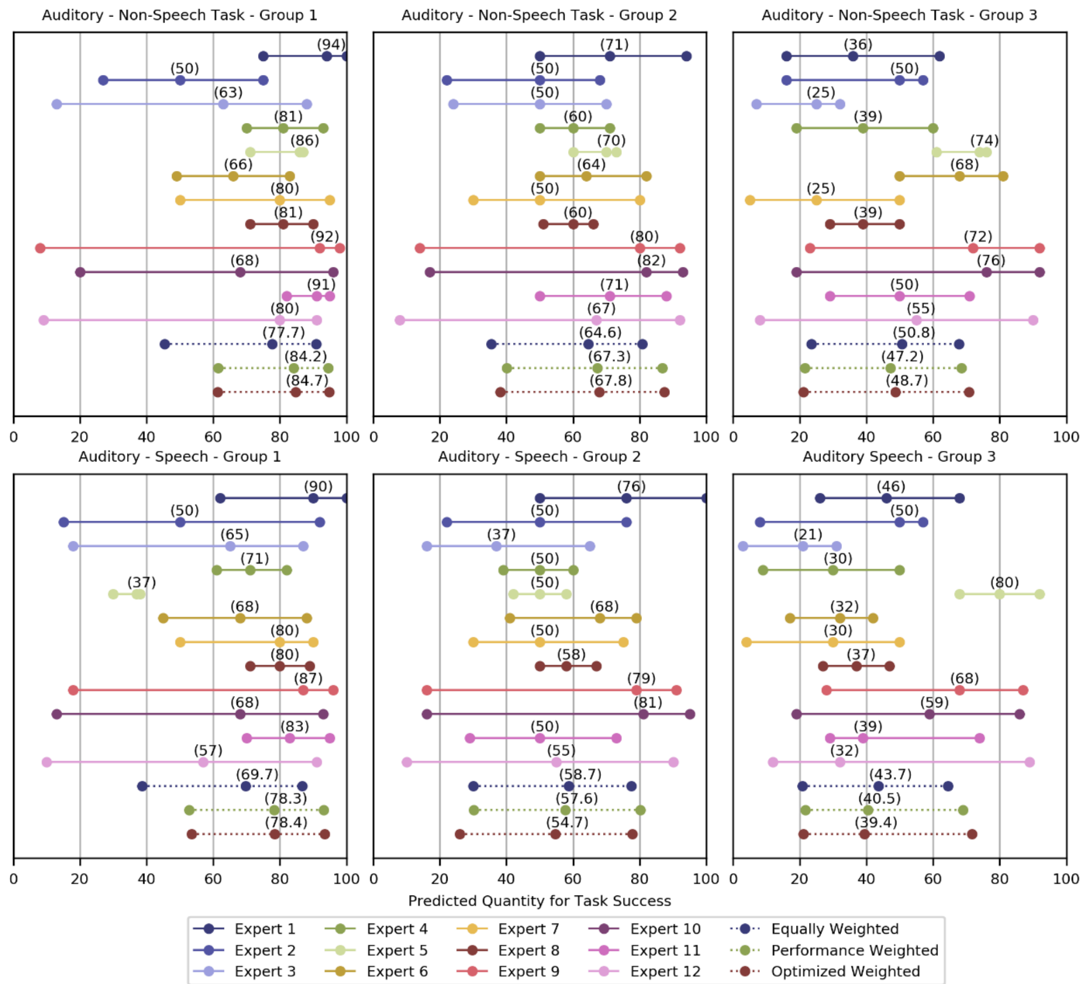


Figure 9: Elicitation results for auditory tasks.

Figure 10 contains results for applying existing knowledge tasks. For these tasks and user groups, risk increased with user group risk-level, however differences between task complexity within risk-levels were less pronounced.

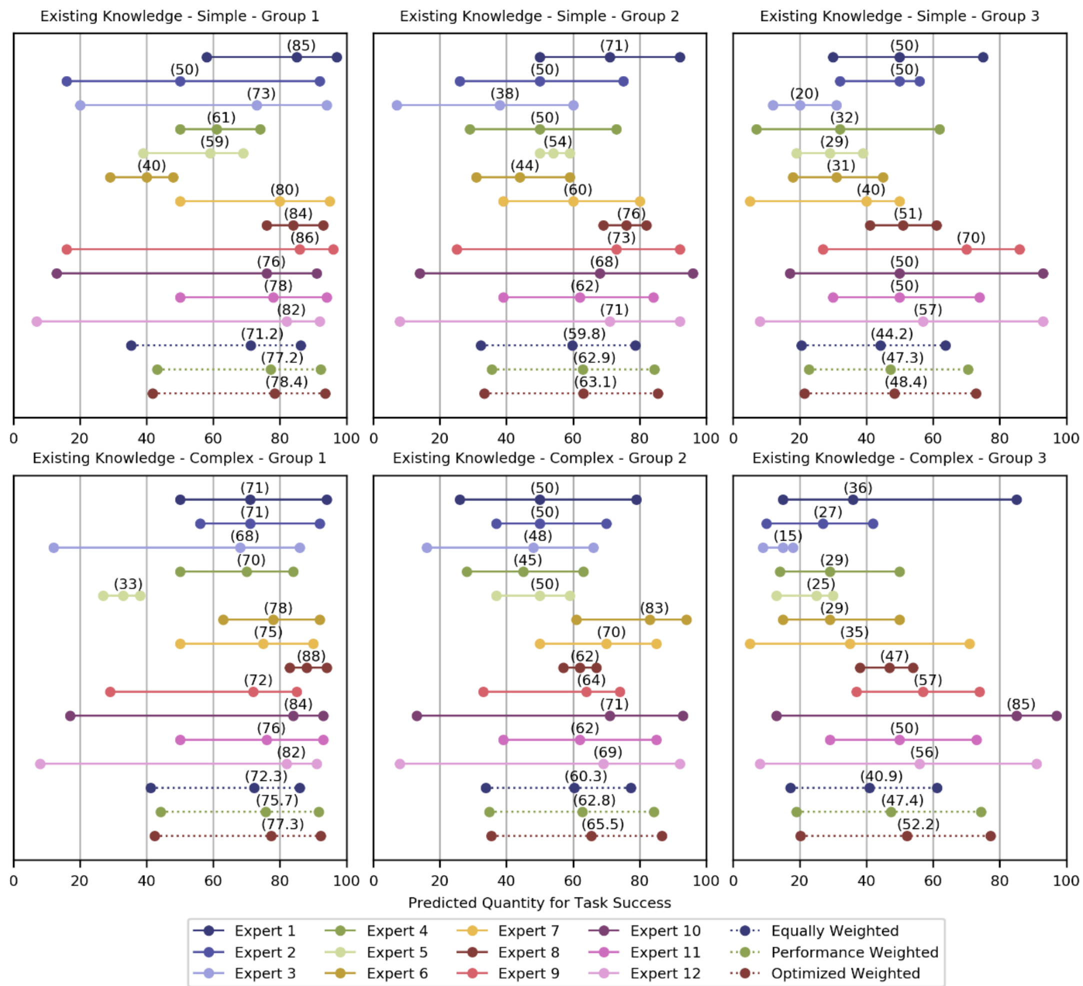


Figure 10: Elicitation results for applying existing knowledge tasks.

Figure 11 contains results for the problem-solving and decision-making tasks. Once again, similar trends can be observed for increasing risk and increasing user group risk-level. Task complexity was only noticeably different between groups 1 and 2, however. User group 3 elicited similar perceived risk across task complexity. Comparing Figures 10 and 11, it can be observed that elicited risk across task complexities were not dramatically different either.

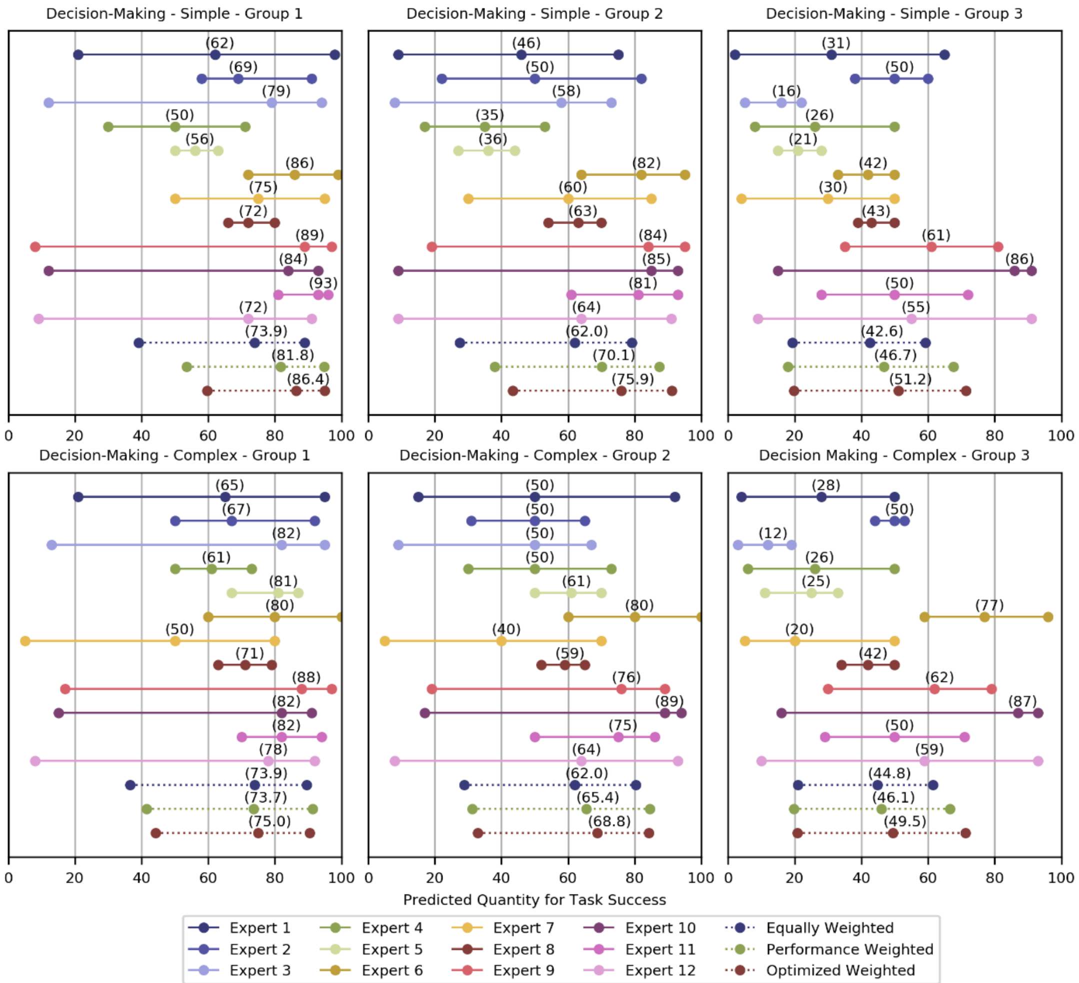


Figure 11: Elicitation results for problem-solving and decision-making tasks.

5.4.4 Model Results

<u>Background Questions</u>: Linear mixed-models were fit to examine the relationships for expected values and information scores for all tasks with background question responses, treating expert ID as a random effect. This was also performed for calibration question scores and background question responses. In all cases, there were no significant associations detected between the elicitation outcomes and the background questions.

Risk-Levels: Expert elicited expected values for task performance were modeled with task risk-level. Model fixed effects are reported in Table 29. Sample size was n = 324. Conditional r^2 was 0.557 and marginal r^2 was 0.232. Random effect standard deviation was 9.79 for Expert ID and 5.25 for task-type. Risk-level 1 is the reference level. Multiple comparisons results using Tukey's method are shown in Table 30. Expect values followed a logical pattern, where increasing risk-level corresponded with decreased likelihood of task success.

Table 29: Linear mixed-model fixed effects for risk-level expected value analysis.

Fixed Effect	Estimate	Std. Error	df	t-value	*p-value*
Intercept	73.3	3.48	18.7	21.04	<.001
Risk-Level 2	-9.65	1.91	308	-5.05	<.001
Risk-Level 3	-21.8	1.68	300	-12.9	<.001

Table 30: Multiple comparisons results for expected value by risk-level analysis using Tukey's method. P-value in parentheses. Difference is row – column.

	Risk-Level 2	Risk-Level 3
Risk-Level 1	9.65 (<0.001)	21.783 (<.001)
Risk-Level 2	-	12.133 (<.001)

Information scores were also modeled with task risk-level. Model fixed effects are reported in Table 31. Sample size was n = 324. Conditional r^2 was 0.502 and marginal r^2 was 0.014. Random effect standard deviation was 0.365 for expert ID and <0.001 for task type. Risk-level 1 was the reference level. Multiple comparisons results using Tukey's method are shown in Table 32. Experts seemed to be the least confident for the second risk level, with only a significant difference in information between risk-level 1 and 2.

Table 31: Linear mixed-model fixed effects for risk-level information score analysis.

Fixed Effect	Estimate	Std. Error	df	t-value	*p-value*
Intercept	0.786	0.111	12.6	7.07	<.001
Risk-Level 2	-0.155	0.053	296	-2.92	0.004
Risk-Level 3	-0.094	0.048	297	-1.97	0.049

Table 32: Multiple comparisons results for information score by risk-level analysis using Tukey's method. P-value in parentheses. Difference is row – column.

	Risk-Level 2	Risk-Level 3
Risk-Level 1	0.1547 (0.009)	0.0939 (0.121)
Risk-Level 2	-	-0.0608 (0.485)

Task-type: Expected values were modeled based on their highest-level task grouping (physical, sensory, cognitive). Model fixed effects are reported in Table 33. Sample size was n = 324. Conditional r^2 was 0.282 and marginal r^2 was 0.037. Random effect standard deviation was 9.59 for expert ID. Cognitive task was the reference level. Multiple comparisons results using Tukey's method are shown in Table 34. Experts elicited larger values for physical tasks than both sensory and cognitive tasks. Information score by task-type was also examined, however no significant associations were identified.

Table 33: Linear mixed-model fixed effects for task-type expected value analysis.

Fixed Effect	Estimate	Std. Error	df	t-value	*p-value*
Intercept	58.9	3.09	13.8	19.1	<.001
Physical	8.86	2.52	309	3.52	<.001
Sensory	1.84	2.37	309	0.778	0.437

Table 34: Multiple comparisons results for expected value by task-type analysis using Tukey's method. P-value in parentheses. Difference is row – column.

	Sensory	Cognitive
Physical	7.02 (0.038)	8.86 (0.001)
Sensory	-	1.84 (0.715)

5.5 Discussion

In this section, the results are interpreted and discussed in the context of reliability and design validation.

5.5.1 Calibration Question Performance

Prior to eliciting values for the unknown quantities, experts were asked to estimate values for five calibrations questions with known values. These questions were related to recent statistics about the diabetes population. Expert performance is summarized in Table 26.

There was a relatively wide distribution of information scores, with some experts opting to provide small confidence intervals while others expressing less confidence. The maximum information score achievable given the range of available answers (0-100) was 3.337, corresponding to percentiles (0,1,2) and (98,99,100). We can see that mean information scores ranged from 2.3% to 34% of this maximum. It is difficult to make a value judgment on the information score alone. Ideally, a larger score is preferable, however, poor information may reflect the nature of the question being asked. Experts were asked to make judgments about very general tasks with few details on the context. It should be expected that there is a high degree of variability in risk per individual even within user groups. Further, several of the experts with high information scores had very poor calibration. This shows that overconfidence can adversely impact statistical accuracy.

Given that there were only five calibration questions, the theoretical maximum calibration score of 1 was impossible to achieve. Five questions cannot be evenly disturbed into elicited intervals as the expected reference proportions:

[0.05,0.45,0.45,0.05]. The maximum calibration score that could be achieved was 0.740, which corresponds to calibration question answers being distributed into the specified intervals as [0,2,3,0] or [0,3,2,0]. There were two experts who received this calibration score, 7 and 12. Both experts were included in the optimized combination of experts. The highest scoring (calibration * information scores) expert was expert 11. In the weighted combination of experts, their estimates were weighted at 43%, approximately 5x their contribution under equal weighting. Expert 11 provided the most balance between information and calibration, demonstrating the importance of balancing confidence with precision when making judgments.

Little difference in performance was observed between calibration questions. Of the five, experts seemed to have the most difficulty with "How many US adults with diagnosed diabetes would you expect to have an A1C value of greater than 9.0%", with only 50% of experts being "correct". This may be because of variation in clinicians' personal experiences with patient care rather than reliance on national aggregate estimates.

Background questions regarding years of experience and experience with at-risk patient populations had no significant association with calibration question performance. This is not a negative, as this provides evidence that physicians of most backgrounds are suitable candidates for this type of elicitation. Future work should seek to validate this finding.

5.5.2 Task Elicitation Results

Elicited values and the combination decision-makers are displayed in Figures 6-11. The optimized weighted decision-maker produced the best results on calibration questions (Table 27). The assumption is, therefore, that this combination of experts should be used in practical application and will be the primary focus of this discussion. On visual inspection, for physical and sensory tasks, risk follows a logical trend. As risk-level and task complexity increases, patient likelihood of success decreases. For cognitive tasks, however, these differences are not as pronounced, especially when looking at similar risk levels across tasks. It may be that the abstract and unobservable nature of cognition makes it difficult to discriminate between various levels of cognitive difficulty.

Comparing the tasks with two patient groups (fine motor movement, visual discrimination) with the three patient group tasks, a potential confounding influence given the ordered presentation of patient risk-levels can be seen. The 2nd group in each of the 2-group tasks seems to resemble the 2nd group in the 3-group tasks, whereas one might expect the 2nd group in the 2-group tasks to be lower, falling somewhere between the level-2 and level-3 risk categories, or potentially lower. It is possible that when participants were taking the survey, they anticipated a 3rd, higher-risk group to be presented and left "space" on the lower end of the scale. This bias may have been mitigated with randomization of questions. Another explanation is that the performance distributions elicited truly represent the predominant risk-levels associated with the diabetes population, and that physicians see less risk involved with visual discrimination and fine motor movements than the other tasks presented.

Statistical modeling was used to further examine the influence of risk-level on estimations (Tables 29-32). For this analysis, the lowest group in each task was coded as risk-level 3, an assumption that may have been incorrect given the previous insights. Regardless, a logical trend was observed, where each risk-level elicited a decreasing expectancy of patient success. Examining the information scores by risk-level, risk-level 1 elicited more confidence from experts than risk-levels 2 or 3. This is likely because higher risk patients present more complicated and heterogeneous use cases, therefore making it more difficult to predict how any one individual will perform a task.

Based on modeling efforts to examine differences between tasks at a high-level (Tables 33-34), physical tasks were perceived as less risk-inducing than sensory and cognitive tasks for the included patient user groups. There was no significant difference between sensory and cognitive tasks. It is possible that the difference between physical and cognitive performance is exaggerated because all cognitive tasks contained a "3rd risk-level", however this does not explain the difference between physical and sensory tasks, which both have one task with only two user groups. It is possible that physicians truly perceive physical tasks to be less risk-inducing within the context of medical device use.

5.5.3 Proposed Use

The goal of this elicitation was to produce values for human performance to support design validation and product customization for products used by highly heterogeneous user populations, where direct measurement is difficult or infeasible. The values elicited in the demonstration were proportions, however this can be easily translated into probabilities. These probability values can be used to predict where user failure is likely to occur during formative human factors design analysis and can be used to justify inclusion of functionality for certain user sub-populations. For example, user groups with high probability of failing to perform visual discrimination tasks may need design features than translate output stimuli to an auditory modality, or vice versa in user populations with high risk of failing to discriminate auditory stimuli (Dascalu et al., 2017; Niazi et al., 2016). Using the user group proportions identified in Chapter 4 (Tables 11,13,15,17, and 19), the magnitude of risk and potential cost to the firm designing the system can be estimated. If cost models were developed for design solutions associated with different tasks, these models could be integrated into an optimization model with the elicited statistics to support cost-effective design decision-making. **A process to accomplish this is proposed and demonstrated in Chapter 6.**

Another advantage of these statistics is that they can be used in early stages of design. Methods for incorporating human factors considerations in early design stages has been identified as a research need (Ozcelik et al., 2011; Sun et al., 2018), particularly for patient populations (Nelson et al., 2016). In early design stages, design choices can be made that may inadvertently disadvantage certain users, yet with no physical product design, specific user interactions have yet to specified. Further, not considering human performance until the late design stages can lead to costly redesign requirements, or certain user's needs being ignored (Leonard et al., 2006). By eliciting performance for generalized tasks, probable interactions can be inferred for a system at early design stages and performance for those interactions can be estimated. While these estimations likely will not be as accurate without knowing the specific interaction, they can be used prior to time consuming and expensive prototyping.

These values could also be used as prior knowledge to support more targeted applications. In Bayesian statistics, Bayes' theorem is used to update probability distributions given new data or information. From this perspective, if some limited performance data was available or was collected to model performance for a specific use case, the values elicited in this study could be used as an informative prior in a Bayesian analysis of the data and be tailored to that specific application (Albert et al., 2012). This provides cost-benefit balance, by providing more confidence in models derived from sparse experimental data. It is not uncommon to use expert elicitation in conjunction with Bayesian inference (N. Wang et al., 2018; Zhang & Thai, 2016).

There were several observed limitations associated with this approach. There was some evidence, discussed previously, that the ordered nature of the questions may have influenced the values elicited from physicians. Future work should examine the extent to which this effect is present for physician expert populations. Another limitation was the decomposition of cognitive tasks. Having now observed that experts were unable to discriminate between cognitive tasks, it may have been better to merge these tasks in favor of adding a wider variety of additional tasks, such as tactile discrimination. Alternatively, additional calibration questions could have been elicited to produce more robust estimations of expert performance.

Chapter 6: Optimizing Function Allocation for Accommodation of Heterogeneous Populations

6.1 Introduction

This chapter introduces a functional modeling approach to facilitate allocation of product functions to humans and machines. This modeling approach is integrated into a multi-objective optimization model to support automated decision-making and to illustrate trade-offs between accessibility and cost. The optimization model automates function allocation for a family of products, where product family members correspond to user groups with varied functional requirements. The tasks generated in Chapter 4 are mapped to functions to facilitate evaluation of function allocations. User groups task performance distributions from Chapter 5 are used to evaluate user accommodation for assigned human functions and are used as input to the optimization model (Figure 12). The output of the methodology are pareto optimal solutions for function allocation.

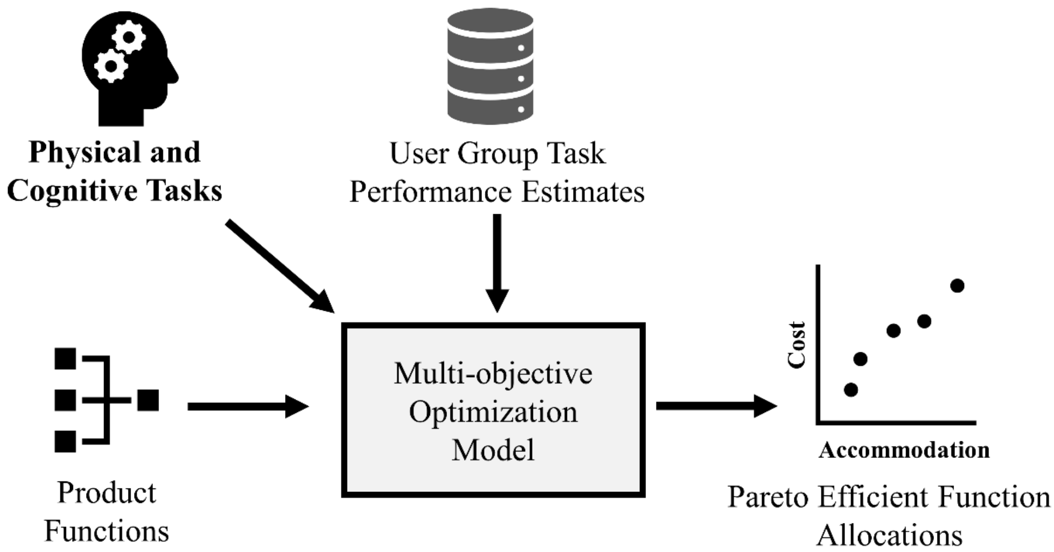


Figure 12: Summary of Chapter 6.

To validate the functional modeling approach and the optimization model, senior mechanical engineering students were recruited to perform three design exercises. The goal of the first two exercises was to compare concepts generated using conventional functional modeling approaches vs. the proposed approach. The goal of the third exercise was to validate the utility of the optimization model.

6.2 Methodology

This section describes the methodology. First, a human-machine system functional modeling approach is proposed, where human and machine functions are represented simultaneously and can be used to facilitate allocation of functions. Next, metrics for evaluation of functional product family concepts are proposed. This includes a theoretical framework for mapping system functions to human tasks to support evaluation of accessibility. Finally, the modeling approach and metrics are integrated into an optimization framework that can produce pareto-efficient design alternatives. The methodology is then demonstrated on the diabetes self-management device case study.

6.2.1 Human-Machine Function Modeling

In this section, a human-machine system perspective to function modeling is introduced. The focus of the modeling approach in this work is to facilitate the allocation of functions. Some of the detail typically included in function structures are not necessary for this task and may be unknowable at the early stages of the design process. Thus, a simplified version of function structure is utilized. Table 35 highlights the differences between the traditional and proposed modeling approaches.

Table 35: Highlighted difference between function structure and proposed model.

Property	**Function Structure**	**Human-Machine Function Model**	**Justification for Difference**
Intended Use	Early (Concept Generation), Mid-Late (Analysis, Redesign)	Early (Concept Generation)	Designers must allocate functions as early as possible to minimize risk of redesign due to discovered incompatibilities between users and functions.
Product Functions	x	x	-
Human Functions		x	Facilitates function allocation between humans and machines.
Flow Directionality	x		Flow directionality provides an unnecessary level of detail for function allocation.
Model Representation	Flowchart	Table	The flowchart is unnecessary without flow directionality.
“Material” and “Signal” as flows	x	x	-
“Energy” as a flow	x		All human functions require energy and is redundant to state.
Level of Detail	Primary and Secondary Functions	Primary Functions	Secondary functions provide an unnecessary level of detail for function allocation and may be unknowable at this early stage.

Modeling Intent: The product of this method, which are a set of functions and their allocation status (human or machine), should be thought of as a precursor to a more detailed function structure where identified product functions are decomposed further. Functions that should be included are those critical to the primary functioning of the system and should be functions that could be conceivably performed by a human or a machine. The distinction between primary and secondary functioning is subjective, and the level of detail to include is ultimately up to the individual performing the method. Primary functions should include functions that the stated problem could not be solved without, and for which there are few or no alternatives. Secondary functions are supportive or accessory in nature. Examples of secondary functioning could include functions required to supply energy to a primary function, functions to facilitate interface between human and product, or functions to describe communication between the system and an entity external to the system (Y. Xu & Huijun, 2007).

Modeling Steps: The first step in most functional modeling approaches and in this modeling approach is to create a black box diagram (Nagel et al., 2015). The black box diagram should state the purpose of the system, as well as key inputs and outputs. Next, the black box should be decomposed into its constituent functions using language from the functional basis, a taxonomy of functions for engineering design (Hirtz et al., 2002). Inputs that each function act upon should be identified as well. The final step is to assign functions to human or machine.

Model Assumptions: There are several assumptions associated with this approach. Allocations are assumed to be static, as opposed to dynamic. While dynamic function allocation is often seen as superior to static because it is more robust to unpredictable conditions, it is outside the scope of this work. The primary goal of this modeling approach is to facilitate accessible design by identifying functions that users cannot reliably perform under any circumstance. This modeling approach also does not consider emergent functions. A one-to-one substitution of human with machine functions can often results in emergent system properties (Dekker & Woods, 2002). For example, if a product is assigned a sensing function then it must also be assumed that a function to communicate sensed information to the user is necessary. A thermometer must sense *and* indicate temperature to be useful. Emergent functions can be identified individually by designers.

6.2.2 Evaluating Function Allocation

The previous section introduced a descriptive model of function and function allocation. In this section, two objective metrics are introduced for evaluating the configuration of allocations across a product family. These metrics are accommodation, or the ability of the user population to perform required functions, and complexity, an approximate measure of cost associated with the design and manufacture of the product family.

6.2.2.1 Accommodation

Accommodation is the ability of a user to perform the tasks associated with a system function. Like a hardware or software element that embodies a product function, a task can embody a human function. Unlike product functions, for which there are theoretically unlimited design solutions, there are limited human tasks that can fulfill a function. Quantifying user accommodation for these tasks can facilitate evaluation of function allocations.

Function-Task Mapping: To help designers identify tasks relevant to performance of a given function, a function-task matrix was developed. This matrix provides theoretical guidance to select tasks required to perform a function. This concept was informed by previous work where authors connected functions to primarily physical tasks (Soria Zurita et al., 2020). In this work, the tasks included were expanded to cover a wider breadth of human activity including elements of sensation (e.g., vision, auditory) and cognition (e.g., decision making). The rows of this matrix contain functions from the functional basis, which is organized in three tiers. A sample of functions from the secondary and tertiary tiers were selected for inclusion. Formal definitions for functions can be found in (Hirtz et al., 2002).

The columns of the matrix are labeled categories of human activity taken from two well validated taxonomies introduced in Chapter 4: Bloom's taxonomy of the cognitive domain (Bloom, 1956) and Harrow's taxonomy of the psychomotor domain (Harrow, 1972). The use of these taxonomies provides a theoretical baseline to identify function-task links for those wishing to reproduce these methods, as well as providing continuity with the methods proposed in Chapter 4 of this dissertation. The matrix can be thought of as a continuation of Table 6 in Chapter 4, providing a direct path from function to P&C Physician Judgment Tasks.

To link the taxonomy levels and the functions, a requirements elicitation using nominal group technique was performed with three engineering Ph.D. students with formal training in cognitive taxonomies and engineering methodology. Participants were all students at the University of Maryland and members of the Hybrid Systems Integration and Simulation Lab. Nominal group technique typically follows a sequence of independent idea generation, round-robin presentation of ideas, group consensus discussion, and voting (Manera et al., 2019). For each function (row), participants were asked to consider the human actions that could fulfill each function. Then, participants identified the taxonomy categories for each action in the matrix. While typically these taxonomies are thought of as nested (e.g. fundamental movements cannot be performed without reflexive movements), experts were instructed to mark the highest applicable level for each taxonomy. This assumes that to quantify performance for any taxonomy level, the influence of lower levels is inherent and cannot be separated from that action and therefore does not need to be explicitly stated. This nested assumption does not apply to Perceptual Abilities and Non-discursive Communication, which are not considered "nested". If the expert believed two levels could be relevant in different situations, they were instructed to mark both. They could also leave the row blank if there was no analogous human activity. This was performed separately for material and signal inputs. Energy was not included (see Table 35).

Following this, the participants convened to discuss results. For each row, participants were randomly selected to defend their selection. A discussion ensued, and if consensus was met, discussion moved to the next row. If not, vote by majority was used. The final function-task matrix can be seen in Tables 36 and 37 for Harrow's and Bloom's taxonomy, respectively.

Table 36: Function-task mapping for Harrow's taxonomy.

	Reflexive Movements	Fundamental Movements	Perceptual Abilities	Physical Abilities	Skilled Movements	Non-Discursive Movements
Separate				M		
Distribute				M		
Transfer				M	M	
Translate				M		
Rotate				M		
Couple				M		
Mix				M		
Actuate				M, S		
Regulate				M	S	
Change				M, S		
Stop				M, S		
Store				M		
Supply				M		
Sense			M, S			
Indicate		M				M, S
Process					M	
Support				M		

M = material, S = signal

Table 37: Function-task mapping for Bloom's taxonomy.

	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Separate			M	S		
Distribute			M, S			
Transfer			M			
Translate			M			
Rotate			M			
Couple			M		S	
Mix			M		M, S	
Actuate				M, S		
Regulate						M, S
Change				M, S		
Stop				M, S		
Store	S		M			
Supply	S		M			
Sense						
Indicate						
Process	S	S	M, S	S	S	S
Support			M			

M = material, S = signal

These matrices can be used to determine relevant human tasks given the functions and inputs being evaluated. These should be tasks whose performance can be quantified and used to evaluate accommodation. The exact tasks must be identified by the designer based on the application. This will vary based on assumptions made about the design context, the availability of data, and the desired specificity of the task. For example, “Perceptual Abilities” could be decomposed into visual and auditory discrimination. Even further, visual discrimination of specific stimuli, such as text, can be specified if it is expected to be particularly relevant and to influence performance significantly. In this work, taxonomy tasks were already mapped to quantifiable tasks (P&C Physician Judgment Tasks) in Chapter 4, Table 6.

Quantifying Task Performance: After relevant tasks have been identified using the function-task matrix, performance for those tasks can be quantified. Accommodation for a single function is formulated as the probability that a user will successfully perform the tasks required for that function. Quantifying this for a population could be accomplished in several ways and selecting the best way for the given use case is outside the scope of this methodology. Options include recruiting and studying live participants, surveying users, soliciting expert judgments (as performed in Chapter 5), or utilizing existing data. What must be accomplished, however, is that users must be assigned a probability of success for every task identified given the modeled functions. A user can be represented as a vector of success probabilities where each entry is associated with a specific task.

Modeling a population of users as such can be accomplished in several ways. To illustrate this, two options are presented. **Option #1:** Stratified Population – The user population is stratified into user groups based on similar task performance. Each user group is assigned a unique probability vector, as well as the proportion of users it represents. **Option #2:** Unique Users – A model population of users is produced, and each user is represented as a vector of probabilities. These probabilities represent the individual's likelihood of succeeding at each task. This is a more realistic option but requires a more complex model.

Accommodation Metric Definition: Given a model user population with accompanying task performance values, accommodation can be evaluated. Accommodation for a product family configuration is measured as the average probability of successfully performing all required functions for all users in the population. For this model, functions exist as independent nodes that can be statically allocated as human or machine. The assumption of independence means that the human performance of one function does influence the subsequent function. Further, the effect of functions performed simultaneously (implying multi-tasking) is not modeled.

Each product in a product family can be represented as a vector of *n* binary values where 1 corresponds to a machine function and 0 corresponds to a human function. A product family *x* is therefore represented as a *m x n* matrix, with *m* product family members. Thus, accommodation for a single individual can be evaluated as follows, where $p = (p_1, \ldots, p_{k=n})$ with entries corresponding to function-task success probabilities, and x_{jk} corresponding to the *kth* function of the *jth* product family member in the product family (Eq. 6):

$$a_j = (a_1, \ldots, a_k)_j = \left(p_1 * [x_{j1} = 0], \ldots, p_k * [x_{jk} = 0]\right) \quad \text{for } j = 1, \ldots, m, k = 1, \ldots, n \tag{6}$$

where square brackets with a conditional statement are the Iverson bracket, evaluated as (Eq. 7):

$$[Z] = \begin{cases} 1 & if\ Z\ is\ true; \\ 0 & otherwise \end{cases} \tag{7}$$

Following this, the product of non-zero elements for each product family member *j* in *a* is taken. It is assumed that a user will prefer the product family member that results in the highest likelihood of success for them. The individual is "assigned" to a product in the family by taking the maximum value in the resulting vector (Eq. 8):

$$A = \max_{j \in J} \left(A_1, \dots, A_j\right) = \max_{j \in J} \prod_{\{k:\, a_{jk} \neq 0\}} a_{jk} \tag{8}$$

The index of the maximum value is retained to allow determination of the distribution of product family member selection in the population.

Example: To demonstrate, suppose two products in a product family have four functions and a user has corresponding success probabilities of p = (0.65, 0.75 0.89 0.98). The product family and function allocations are (Eq. 9):

$$x = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{9}$$

where each row corresponds to a product family member. Accommodation is calculated as follows (Eq. 10-11)

$$a = \begin{bmatrix} 0.65 * 1 & 0.75 * 0 & 0.89 * 0 & 0.98 * 1 \\ 0.65 * 1 & 0.75 * 0 & 0.89 * 1 & 0.98 * 1 \end{bmatrix} = \begin{bmatrix} 0.65 & 0 & 0 & 0.98 \\ 0.65 & 0 & 0.89 & 0.98 \end{bmatrix} \quad (10)$$

$$A = \max\left(\begin{bmatrix} 0.65 * 0.98 \\ 0.65 * 0.89 * 0.98 \end{bmatrix}\right) = \max\left(\begin{bmatrix} 0.637 \\ 0.567 \end{bmatrix}\right) = 0.637 \quad (11)$$

Thus, the user was assigned to product family member 1. For a population of users $i \in I$, accommodation is determined as (Eq. 12):

$$A_{Population} = \bar{A}_i \quad (12)$$

6.2.2.2 Complexity

Measuring cost at this stage of design is difficult because a significant number of design decisions are yet to be made. Yet, with a few assumptions, an approximation of cost can be made that permits comparison between product family configurations. This metric is loosely based on the work in (Gill et al., 2017), where machine learning was demonstrated as useful in some cases for predicting product price from function structures.

Complexity is split into two sub-metrics: Machine Function Cost (MFC) and Unique Function Cost (UFC). MFC seeks to approximate the production cost associated with each non-human function produced. UFC seeks to approximate the upfront design and manufacturing costs associated with producing each unique function.

MFC Metric Definition: MFC is evaluated as the average number of machine functions per individual in the population. This is calculated as follows (Eq. 13):

$$C_{MFC} = \frac{\sum_{j=1}^{m} \sum_{k=1}^{n} x_{jk} * u_j}{U} \tag{13}$$

where u_j is the number of users assigned to product family member *j*, and *U* is the number of users in the population.

UFC Metric Definition: UFC attempts to approximate the costs associated with design and manufacturing of unique functions. The metric is evaluated as the number of unique automated functions across all product family members (Eq. 14).

$$C_{UFC} = \sum_{k=1}^{n} \left[\sum_{j=1}^{m} x_{jk} \geq 1 \right] \tag{14}$$

Each column corresponding to a function in the product family matrix is summed, and if the value is greater than 1, then a cost of 1 is incurred. Thus, holding all other objectives constant, the cost metric will prioritize solutions that minimize the total number of unique machine functions and will promote sharing across product family members. Total complexity is formulated as (Eq. 15):

$$C_{tot} = \alpha * C_{MFC} + (1 - \alpha) * C_{UFC} \tag{15}$$

where alpha is a weight parameter (0-1) that can be tuned given the relative importance of each metric.

Example: Using the same product family configuration as in the prior example, the complexity metric is demonstrated. A population of 100 users is assumed, with 25 assigned to product family member 1 and 75 to number 2. Equal weighting is assumed ($\alpha = 0.5$).

$$C_{MFC} = \frac{(2 * 25) + (1 * 75)}{100} = 1.25 \tag{16}$$

$$C_{UFC} = 0 + 1 + 1 + 0 = 2 \tag{17}$$

$$C_{tot} = 0.5 * 1.25 + 0.5 * 2 = 1.625 \tag{18}$$

6.2.3 Optimization Model

Given the objective metrics for accommodation and complexity, a trade-off emerges. To accommodate the needs of more users, more functions must be automated. Some of these costs can be minimized through product family design. A multi-objective optimization model, using the previous metrics as objectives, was formulated to aid designers in navigating these trade-offs (Eq. 19).

$$
\begin{gathered}
\min_{x} F(x) = \{-f_1, f_2\} \\
f_1(x) = A_{Population} \\
f_2(x) = C_{tot} \\
s.t.\ x_i \in \{0,1\}, \qquad i = 1, \dots, m * n
\end{gathered}
\tag{19}
$$

The decision-variable *x* is constrained to whole values of 0 or 1, with the number of variables determined by the number of functions in each product and the maximum number of product family members included. The first objective function is made to be negative so that it can be minimized instead of maximized to conform with typical notation.

Solver Selection: The resulting optimizing problem is discontinuous, non-convex, and non-linear, which is difficult to solve using gradient-based approaches. As such, a genetic algorithm (GA) is used. GAs are a heuristic approach to optimization problems that don't rely on gradients and are therefore well suited for this problem (Q. Wang et al., 2019). GA's perform a semi-random search in the objective space. Candidate solutions are represented as chromosomes, where genes belonging to chromosomes represent each decision-variable. Chromosomes are randomly generated and evaluated on a fitness function (comprised of the objective functions). Chromosomes are ranked based on fitness and a proportion of the chromosomes with the highest fitness are kept. These chromosomes (parents) undergo two mechanisms (cross-over, mutation) to generate children chromosomes, which are then in turn evaluated in the objective space (Alizadeh et al., 2019). This continues until some cut-off criteria is satisfied. A disadvantage of this type of solver is that there is no guarantee the identified solutions are global optimum, which may not be a problem if local optima are satisfactory (J. Wu & Azarm, 2000). If the function model is particularly large, and solver run time becomes intractable, it may be desirable to use a surrogate model to simplify the problem (Chatterjee et al., 2019).

Unlike single objective optimization, in multi-objective optimization problems a single solution is typically not obtained. Instead, a set of solutions is obtained that are, ideally, pareto efficient. Pareto efficient refers to a solution that cannot be improved in one objective without worsening another. The set of solutions that are pareto efficient is referred to as a pareto front (M. Li et al., 2020). This solution-set is the final output of this optimization model. Providing this solution set allows stakeholders to examine how design decisions influence objective function trade-offs and can facilitate selection of a solution based on stakeholder values.

6.3 Experimental Validation

To demonstrate the proposed function modeling approach and validate the utility of the optimization model, senior mechanical engineering capstone design students at the University of Maryland were recruited (n=16) to perform three design exercises. These students were all enrolled in the course ENME 472: Integrated Product and Process Development. The first two exercises were developed to demonstrate the proposed human-machine function modeling approach (exercise 2) in comparison to the conventional approach (exercise 1). In these exercises, the modeling approaches were applied to a mobility assist device design case study. The third exercise was developed to validate the utility of the optimization model. Participants were required to allocate pre-selected functions for the diabetes self-management device case study. The tasks and purpose for each student exercise are summarized in Table 38 and discussed further in the following sections.

Table 38: Summary of student exercises.

	Human-Machine Function Modeling Validation		**Optimization Model Validation**
Title	Conventional Modeling Approach (Exercise 1)	Proposed Modeling Approach (Exercise 2)	Manual Function Allocation (Exercise 3)
Summary	Participants perform <u>conventional</u> modeling approach on design case study	Participants perform <u>proposed</u> modeling approach on design case study	Participants allocate functions to humans and machines for a design case study
Select functions?	X	X	
Allocate functions?		X	X
Applied to…	Mobility assist device case study	Mobility assist device case study	Diabetes self-management device case study
Compared to…	Exercise 2	Exercise 1	Optimization model output

Due to the COVID-19 pandemic, which occurred during the time of this data collection, exercises were performed remotely. The exercises were implemented using Qualtrics and went through several rounds of internal piloting before dissemination to participants. Dissemination occurred in Spring 2020. The students were incentivized with 2 points of extra credit towards their final homework grade (0.05% points on the final grade). This project received IRB approval after an expedited review as a minimal risk project (IRB Package # 1559396-2).

6.3.1 Human-Machine Function Modeling Validation

The goal of the first two exercises was to investigate the influence of the proposed function modeling approach on concept development compared to conventional, product-centric approaches. Prior to the exercises, participants were presented with a design case study (<u>unrelated</u> to the diabetes self-management device case study) where they were asked to generate a concept for a family of mobility assist devices. A different case study was generated to prevent learning effects on the subsequent Exercise 3, which was applied to the diabetes self-management device case study. The mobility-impaired population was selected for this case study because it is a broad population with a variety of additional usability needs. Further, mobility devices are primarily physical and are therefore well suited for undergraduate mechanical engineering students. Lastly, statistics were readily available for this population in NHANES.

Participants were given case study material that contained a problem description, as well as summary statistics generated for the population using 2017-2018 NHANES data. The problem description was as follows:

> "*The most common disabilities in the United States are mobility related. Many devices exist to aid mobility, however due to the highly varied characteristics and needs of mobility impaired individuals, not one size fits all. For this task, you will identify functions for a product to aid mobility in individuals with serious difficulty walking. This device should be for usable in, but not limited to, the home and outside on paved surfaces. You will complete two different exercises concerning this design problem. You should try to balance product accessibility with product complexity. Summary statistics for the mobility-impaired population have been provided to help you perform this task.*"

Statistics were generated from NHANES participants who reported "serious difficulty walking or climbing stairs." Variables that were summarized were those included in the Chapter 4, as these were identified as important characteristics for product interaction. The objective of these materials was to provide participants with information they could easily access independently without the use of the methodology proposed in this dissertation. Case study materials as given to participants are contained in Appendix D.1. At the beginning of the Qualtrics survey, participants were also given a brief background on modeling function, product families, and accessible design. To ensure participants read the material, questions pertaining to the material were periodically presented throughout the study. Participants were unable to advance without a correct response to the questions.

<u>Conventional Modeling Approach Exercise</u>: The conventional modeling approach exercise was derived from methods taught in the undergraduate course participants were recruited from. The methods in this course generally adopt a product-centric perspective to modeling product function. Students identify the functions a product performs and the flows (material, energy, signal) they act on. Typically, they will then organize these functions as a block diagram "function structure". Function allocation is not a concept that is introduced in this course.

For the exercise, participants were presented with several tables that contained two drop-down boxes. Drop-down boxes in column one contained functions and drop-down boxes in column two contained generic inputs to the functions (human body, status signal, control signal, object, liquid, gas). Functions were structured as a statement and included the functions from Tables 36-37. Each table corresponded to a different product family member. Participants were instructed to select functions and inputs for a family of products to satisfy the needs of the user population given the provided case study. Participants were required to include at least two product family members with at least two functions, with a maximum of five product family members. Participants were also instructed that products must all be different. Figure 13 shows an example of the function selection tables in this exercise.

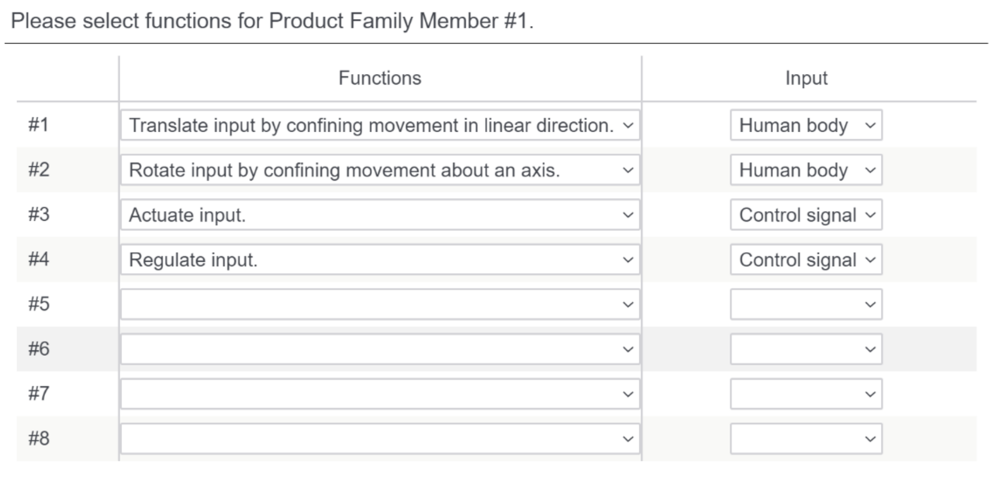


Figure 13: Function selection table example for the conventional approach exercise.

Following this, participants were presented with several open-ended questions. These questions asked participants to generate and describe a product family concept given the selected functions. Participants were able to review their previous selections while answering questions. Questions included the following:

1. Describe the core ideas of the concept at a high-level. You should describe the common features between products and the features that vary for each product (3 or more sentences):

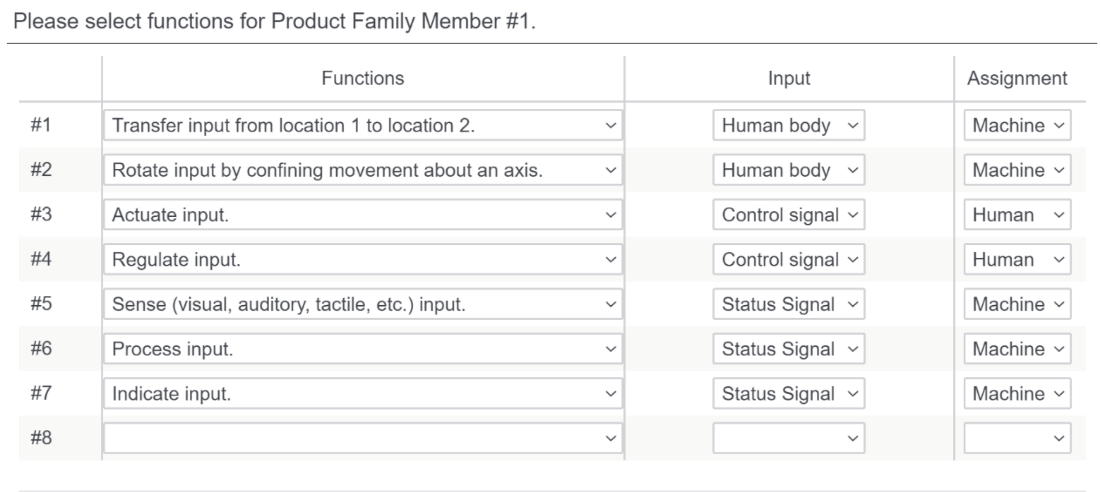
2. Describe the intended user population for each product family member. This should include the characteristics that drove the need for that unique product.
    i. Product family member 1 (1-3 sentences):
    ii. Product family member 2 (1-3 sentences):
    iii. …
3. How useful did you find the case study summary statistics in identifying product family user populations?
    i. Not useful
    ii. Somewhat useful
    iii. Very useful
4. Why did you find these statistics useful or not useful? (1-3 sentences):

<u>Proposed Modeling Approach Exercise:</u> On completion of these questions, the second exercise was presented corresponding to the proposed (human-machine) function modeling approach. The proposed modeling exercise was similar to the conventional modeling exercise, with the additional task of allocating functions to human or machine. Further, this exercise introduced three required functions corresponding to human information processing that occurs in all human-machine systems where the human receives sensory feedback from the system, processes this feedback in the working memory, and stores this information either temporarily or in long-term memory. The functions were stated as:

1. FUNCTION: Sense the state of system functions. INPUT: Function Status Signal
2. FUNCTION: Process sensory information. INPUT: Sensory information
3. FUNCTION: Store sensory information. INPUT: Processed sensory information

Sensing and processing information about the state of a system is a critical task for all products with a human component. These functions were included automatically because it was assumed that the participant population (undergraduate mechanical engineering students) would have little experience in human factors, human-computer interaction, cognitive science, or related disciplines. While this does introduce some bias in the research design, it is unlikely that student participants would naturally incorporate these critical functions in their responses otherwise.

Participants were instructed that they would perform a similar task for the same case study, but this time they would be selecting *who* (human, machine) performs each function in the system, in addition to selecting *what* functions the system performs. Participants were presented with background on function allocation as well as the information processing functions concept. Again, participants were required to define at least two product family members with at least two functions (in addition to the information processing functions), with a maximum of five product family members. Participants were instructed that all product family members must be different, and that selections did not need to mirror those in the prior exercise. Figure 14 shows an example of the function selection tables in this exercise.



Please select functions for Product Family Member #1.

	Functions	Input	Assignment
#1	Transfer input from location 1 to location 2.	Human body	Machine
#2	Rotate input by confining movement about an axis.	Human body	Machine
#3	Actuate input.	Control signal	Human
#4	Regulate input.	Control signal	Human
#5	Sense (visual, auditory, tactile, etc.) input.	Status Signal	Machine
#6	Process input.	Status Signal	Machine
#7	Indicate input.	Status Signal	Machine
#8			

Figure 14: Function selection table example for proposed approach exercise.

Participants were then presented with the same questions as in the first exercise to facilitate comparison. Table 39 provides a comparison of the conventional and proposed exercises.

Table 39: Comparison of convention model approach exercise and proposed model approach exercise.

	Conventional Modeling Approach Exercise	**Proposed Modeling Approach Exercise**
Participant selects functions.	x	x
Participant selects inputs.	x	x
Participant allocates functions.		x
Includes information processing functions.		x
Participant defines up to 5 product family members.	x	x
Participant generates a text-based concept.	x	x

<u>Analyzing Free Response Questions</u>: To evaluate free response questions for the conventional and proposed modeling exercises, qualitative coding was utilized. Qualitative coding is the process of assigning categories to passages of text based on prevalent themes to infer trends (Castleberry & Nolen, 2018). Coding was performed inductively, where codes were generated while examining the text, as opposed to

forming codes prior to reading the text (deductive) (Kalpokaite & Radivojevic, 2019). Coding was performed by two researchers at the Hybrid-Systems Integration and Simulation Laboratory at the University of Maryland. The process began with researchers independently reading through the text and generating a broad list of potential thematic codes. To guide the code generation process, categories were generated prior. The categories corresponded to the subject matter of the questions introduced in Chapter 6.3.1 and were generated during a brainstorming session including both researchers. These included: **Concept Features**, **User Population Description**, **Accessibility Requirements**, and **Use of Statistics**. Next, the researchers convened and compiled codes into a final code list where redundant and conflicting codes were removed or combined. Next, researchers returned to the text and applied codes to each question independently. Questions could have more than one code applied if necessary. This included all questions for each exercise. Finally, the researchers reconvened and compared coding results. The researchers discussed any question where there was discrepancy in the applied codes and determined a final coding.

To analyze the codes, codes were collapsed by participant and by exercise. For each participant-exercise pair (n=32), each code was designated as 1 if it was applied at least once, and 0 otherwise. Code and function frequencies were tallied and compared across exercises.

Further, to investigate if participants were using the selected functions to inform their concepts, several research questions were generated regarding function-code co-occurrence for concept feature codes. These research questions were based on logical functions that should have been included given the resulting generated codes. For each participant, a co-occurrence was tabulated if the participant had a code applied to them and they used the function at least once. This was repeated for both exercises (n=32). The same was repeated each time a function *was not* used at least once. Then, these values were turned into probabilities for each function-code pair – Pr(Function = Yes | Code = Yes) and Pr(Function = No | Code = Yes). Finally, these probabilities were used to calculate the odds of a code being applied when a certain function was used. In cases where either of these probabilities was 0, the Haldane correction was applied, where 0.5 is added to the numerator and denominator. Note that, due to the very small sample size, these were only summary statistics and are not useful for hypothesis testing. These values should not be treated as statistically valid and should only be used for hypothesis generating. Further, the large quantity of statistical tests required to hypothesis test function – code co-occurrences would render statistical power negligible.

6.3.2 Optimization Model Validation

To validate the optimization model, participants performed a third design exercise where they were asked to manually allocate functions for a family of products for the diabetes self-management device case study. Participants were tasked with performing the same task as the optimization model by selecting values for the same decision-variables. This enabled direct comparison between human judgment and model performance on the allocation task. Unlike the prior exercises, participants were constrained to a pre-selected set of functions. Otherwise, direct comparison with the optimization model would be difficult as the model only evaluates the allocation of functions, not the selection of functions to include.

First, the application of the optimization model is detailed. Sub-sections follow the order of the methodology discussed in Chapter 6.2, applied to the diabetes self-management device case study referenced throughout the dissertation. This includes discussion on generating a human-machine function model, mapping of those functions to physical and cognitive tasks, quantifying performance for those tasks using values from Chapter 5, developing a model population of users for input into the model, and details on execution of the model. Then, the details of the student manual allocation optimization exercise are discussed.

Case Study Function Model: In most cases, blood glucose monitoring systems work by extracting some glucose carrying medium from the body. This medium is most often blood but can also be urine or saliva. The medium undergoes some change (e.g., a chemical reaction), and from that reaction the concentration of glucose can be derived. This value can then be used to determine if the user is in a healthy state. Table 40 contains functions and function inputs identified for the diabetes blood glucose monitoring system. The first three functions correspond to monitoring of the system by the user, the following seven describe the core functionality of the system. Note that while “condition” and “detect” are not included in Tables 36-37, they are cited synonyms (Hirtz et al., 2002) of “change” and “sense”, respectively, adapted for the specific case.

Table 40: Function model for diabetes blood glucose monitoring system.

Function	Input	Class
Sense	State of system functions	Signal
Process	Sensory information	Material
Store	Sensory information	Material
Extract	Blood glucose from body	Material
Transfer	Blood glucose	Material
Store	Blood glucose	Material
Condition	Blood glucose	Material
Indicate	Glucose levels	Signal
Detect	Glucose levels	Signal
Process	Glucose levels	Signal

Function-Task Mapping: The next step was to map functions to tasks for which performance has been quantified. The mapping was facilitated using Tables 36-37. The specific tasks to be used for the optimization model were the case study tasks identified in Chapter 4, Table 6 and further refined in Chapter 5, Table 23. Table 41 contains the Function – Task mappings.

Table 41: Case study function-task mapping.

Function	P&C Taxonomy Task	P&C Case Study Task
Sense Signal	Perceptual Ability	Visual Discrimination - Simple
Process Signal	Synthesis	Problem-Solving and Decision-Making - Complex
Store Signal	Knowledge	Applying Existing Knowledge - Simple
Extract Material	Physical Ability, Application	Fine Motor Movement, Applying Existing Knowledge - Complex
Transfer Material	Physical Ability, Application	Gross Upper-body Movement, Applying Existing Knowledge - Complex
Store Material	Physical Ability, Application	Gross Upper-body Movement, Applying Existing Knowledge - Complex
Condition Material	None	None
Indicate Signal	None	None
Detect Signal	Perceptual Ability	Visual Discrimination – Complex OR Auditory Discrimination – Speech
Process Signal	Evaluation	Problem-Solving and Decision-Making - Complex

It was determined that conditioning of blood glucose and indication of blood glucose levels would not be feasible actions for a human to perform. For "Sense Signal", where the user is monitoring the state of each system function, it was determined that this would practically require visual discrimination. Whereas "Detect Signal", where the user is detecting output blood glucose levels, could practically be visual or auditory discrimination depending on how the indicate function was fulfilled. Automating this detect function would indicate the need for some alternative means for delivering information to the user.

Task Performance Quantities: Expert estimates from Chapter 5 were used as input to the optimization model. Optimized weighted estimates were used and are displayed in Table 42 with group # corresponding to increasing patient risk.

Table 42: Expert elicited task performance values.

P&C Case Study Task	**Group #**	**5th Percentile (%)**	**50th Percentile (%)**	**95th Percentile (%)**
Gross Upper-body Movement	**1**	60.15	85.06	95.66
	2	47.04	75.17	87.09
	3	45.14	64.79	84.41
Fine Motor Movement	**1**	51.62	87.01	95.26
	2	56.57	80.69	92.94
Visual Discrimination – Simple	**1**	68.01	87.34	95.87
	2	55.51	80.19	91.59
Visual Discrimination – Complex	**1**	45.82	72.33	89.14
	2	31.52	59.83	82.17
Auditory Discrimination – Speech	**1**	53.51	78.41	93.38
	2	26.03	54.69	77.72
	3	21.11	39.40	71.69
Applying Existing Knowledge – Simple	**1**	41.74	78.43	93.65
	2	33.28	63.06	85.41
	3	21.33	48.39	72.96
Applying Existing Knowledge – Complex	**1**	42.37	77.34	92.22
	2	35.46	65.45	86.62
	3	20.20	52.16	77.23
Problem-solving and Decision – Complex	**1**	44.16	74.97	90.49
	2	32.89	68.79	84.25
	3	20.76	49.50	71.23

Population Model: To model the population for input into Equation 6, NHANES participants clustered in Chapter 4 were assigned success probability values from Table 42 based on cluster membership. 50^{th} percentile values were used. NHANES participants are not required to answer every question in the NHANES survey. Therefore, not all NHANES participants were included in all task clusters due to lack of responses for task-specific variables. For example, as shown in Chapter 4, the sample size for gross upper-body movement clusters was 720, while for cognitive task clusters it was 616. Only participants that responded to all NHANES variables across all tasks were included in the population model (n = 616) to ensure that accommodation could be evaluated for all individuals. Task success probabilities were then extracted based on the function-task mappings in Table 41, resulting in a vector of 8 probability values per population member. The max value for auditory or visual discrimination was extracted for "Detect Signal". Functions with more than one associated task received the joint probability of success for tasks.

Optimization Model Implementation: A multi-objective optimization model with a genetic algorithm was implemented in MATLAB using the global optimization toolbox (*MATLAB Ver. R2020b*, 2020). The model was formulated as in Equation 19 with the goal of optimizing function allocations for functions in Table 40. "Condition blood glucose" and "Indicate glucose levels" were determined to be mandatory machine functions and were removed from the optimization problem. The number of product family members was limited to five, due to practical design feasibility constraints. The decision-variable was therefore a 5x8 matrix of values constrained to 0 or 1. The genetic algorithm population size was evaluated in increasing increments of 500 from 500 to 8000, with hyper-area difference (HD) (J. Wu & Azarm, 2000) between the dominated space and the objective space, the number of unique solutions (NS), and solution spread (SS) being used as selection criteria. Spread as described in (Deb, 2009) is utilized, where a lower value indicates solutions are more evenly distributed in the objective space. Maximum generation was set to 300. The algorithm was also set to terminate if the spread of pareto optimal solutions did not improve over 50 generations with a tolerance of 0.0001. The default crossover fraction of 0.8 was used.

Student Manual Allocation Optimization Exercise: In this exercise, participants were asked to manually perform the same task as the optimization model – allocating functions for a family of products. This exercise was also implemented in Qualtrics. Participants were first presented with a brief description of the diabetes self-management case study and population summary statistics for the diabetes population generated from the NHANES data. Again, the objective of these materials was to provide participants with information they could easily access independently without the use of this methodology, thus validating the steps taken in the methodology to produce new information about the population. Participants were expected to use this information to estimate dominant user group characteristics and evaluate their ability to perform product functions. Case study material presented to participants can be found in Appendix D.2.

After reviewing the material, participants were presented with the functions from Table 40. They were requested to select the number of products to include in their product family (up to five) and the allocation of functions for each product with the goal of maximizing accessibility and minimizing cost. Participants were also asked a free response question: “What logic did you use when selecting human or machine assignments for each function?”. Responses were coded as described in Chapter 6.3.1.

Participant and Optimization Model Comparison: Participant manual allocations were evaluated using the optimization model metrics for accommodation and complexity introduced in Chapter 6.2.2. Participant responses could then be compared to the output of the optimization model directly. Responses were plotted together and visually compared. A participant was said to outperform the model if they produced a response that was more pareto efficient than at least one model solution.

6.5 Results

This section summarizes the optimization model output and the student exercises. First, the conventional vs. proposed function modeling exercise results are presented. Then, a descriptive summary of the population model used in the optimization model is presented. Last, the optimization model is demonstrated and compared to the results of manual allocation optimization exercise.

6.5.1 Conventional vs. Proposed Function Modeling Results

This section presents a comparison of the qualitative coding results for the conventional and proposed functional modeling exercises. A separate table will be presented for each category of code. 16 participants completed both exercises. Table 43 contains codes for Concept Features.

Table 43: Concept features qualitative coding results for the conventional (CN) and proposed (PR) function modeling exercises.

Code	Description	CN Count	CN %	PR Count	PR %
Mobility Support	Directly serves to move the user or support movement.	15	0.938	14	0.875
Auxiliary Support	Provides supportive functioning that does not directly aid in moving the user.	5	0.313	3	0.188
Fully Manual	Product is manually powered by the user.	4	0.250	2	0.125
Physical Actuation of Movement	Product requires the user to actuate electronic or other powered components.	4	0.250	7	0.438
Autonomous Movement	Product has features that provide some amount of autonomous control.	1	0.063	2	0.125
Provides Visual/Auditory/Tactile Feedback	Product provides feedback for some sensory modality.	2	0.125	0	0
Wearable/Exoskeleton	Product is wearable or described as an exoskeleton.	3	0.188	2	0.125
Motorized Vehicle	Product is any vehicle that is motorized.	2	0.125	2	0.125
Wheelchair	Product is a wheelchair (motorized or not).	1	0.063	2	0.125
Supportive Object	Product described provides static support to the user (e.g. a cane).	5	0.313	2	0.125
Computer	Product contains a computational element.	3	0.188	4	0.250
Sensor	Product utilizes sensors (for the user or the environment).	7	0.438	8	0.500
Feature for Comorbidity Management	Product contains a feature specifically for management of comorbidities.	4	0.250	3	0.188
Storage Features	Product contains a feature for storage of user belongings.	1	0.063	0	0
Safety/Protective Features	Product provides fail-safe or protective features.	3	0.188	1	0.063

Coding revealed few differences between the exercises, with the largest being observed for “Physical Actuation of Movement” and “Supportive Object”. Table 44 contains codes for User Population Description.

Table 44: User population description qualitative coding results for the conventional (CN) and proposed (PR) function modeling exercises.

Code	Description	CN Count	CN %	PR Count	PR %
Physical Functioning	User described in terms of physical actions they have difficulty performing.	12	0.750	10	0.625
Sensory Functioning	User described in terms of sensory/perceptual actions they have difficulty performing	2	0.125	2	0.125
Disease or Medical Condition	User described by disease or medical conditions they have.	5	0.313	8	0.500
Need, Desire, or Preference	User described by expressed needs, desires, or preferences.	3	0.188	0	0
Weight	User described by weight.	1	0.063	1	0.063
Age	User described by age.	4	0.250	3	0.188
Injury	User described in terms of an injury they suffered.	1	0.063	1	0.063
Activity	User described in terms of their typical activity levels (sedentary vs. active)	1	0.063	1	0.063
Financial Status	User described in terms of financial status/wealth.	1	0.063	0	0
Occupation	User described by their occupation.	2	0.125	1	0.063

The most common way to describe the end-user population was in terms of physical tasks they could not perform or have difficulty performing. Once again, most codes trended similarly for each exercise. Table 45 contains codes for Accessibility Requirements.

Table 45: Accessibility requirements qualitative coding results for the conventional (CN) and proposed (PR) function modeling exercises.

Code	Description	CN Count	CN %	PR Count	PR %
Physical	Concept includes at least 1 feature that caters to the physical needs of the intender user population.	15	0.938	15	0.938
Sensory	Concept includes at least 1 feature that caters to the sensory/perceptual needs of the intender user population.	3	0.188	3	0.188
Cognitive	Concept includes at least 1 feature that caters to the cognitive needs of the intender user population.	3	0.188	3	0.188

Physical accessibility was the most common accessibility requirement to be addressed. There was no difference in frequencies of codes between exercises. Table 46 contains codes for Use of Statistics.

Table 46: Use of statistics qualitative coding results for the conventional (CN) and proposed (PR) function modeling exercises.

Code	Description	CN Count	CN %	PR Count	PR %
Negative					
Statistics too Broad	Statistics are too broad to be useful for the specific application.	1	0.0625	1	0.063
Too Much Information	Case studies provided an overwhelming amount of information.	2	0.125	0	0
No Relationships between Variables	Statistics are difficult to use in isolations. Correlation or co-occurrence between variables is needed.	2	0.125	1	0.063
No Link to Design Decisions	Isolated statistics aren't helpful without knowing how they influence user performance/interaction.	3	0.1875	1	0.063
Positive					
Magnitude of Risk	Statistics helped to clarify the magnitude of the users risks.	4	0.25	1	0.063
Size of Market	Statistics helped to clarify the size of the market for the product.	2	0.125	0	0
Problem Comprehension	Statistics helped to generally clarify the problem.	2	0.125	1	0.063
Design Feature Generation	Statistics helped to inspire or generate design solutions.	4	0.25	3	0.188
Population Needs/Struggles	Statistics helped to clarify the specific needs of the user population.	10	0.625	10	0.625
Interesting	Statistics were interesting.	0	0	2	0.125

Participants most commonly cited the statistics as being useful for generally understanding the needs of the user population. On the negative end, several participants said that the statistics were not useful without knowing how they should influence their design decision-making.

Function Frequency: Table 47 contains frequency of function use by participants by exercise. Functions were counted once for each participant. Functions did not differ much between exercises. Transfer was one of the most commonly used functions, along with actuate, sense, and process.

Table 47: Function usage frequency for the conventional (CN) and proposed (PR) function modeling exercises.

Function	**CN Count**	**PR Count**
Separate	5	6
Distribute	6	5
Transfer	16	9
Translate	10	8
Rotate	10	8
Couple	9	8
Actuate	11	14
Regulate	7	9
Change	7	7
Stop	6	8
Store	9	6
Supply	6	5
Sense	13	14
Indicate	6	7
Process	10	11
Support	9	9

Code-Function Co-occurrence: To investigate if participants were using the selected functions to inform their concepts, several research questions were generated regarding function-code co-occurrence for concept feature codes. These research questions were based on logical functions that should have been included given a certain code. Hypotheses generated were:

1. Were codes related to movement (Mobility Support, Fully Manual, Physical Actuation of Movement, Autonomous Movement) more likely to be applied when functions related to movement (Transfer, Translate, Rotate) were used?

2. Was the "Sensor" code more likely to be applied when the sense function was used?
3. Was the "Supportive Object" code more likely to be applied when the support function was used?
4. Was the "Computer" code more likely to be applied when the process function was used?
5. Was the "Physical Actuation" code more likely to be applied when functions related to control (Activate, Regulate, Change, Stop) were used?
6. Was the "Provides Visual/Auditory/Tactile Feedback" code more likely to be applied when the indicate function was used?

Table 48 addresses research question 1.

Table 48: Co-occurrence odds for movement-related functions with movement-related codes.

	Mobility Support (n=8)	Fully Manual (n=6)	Physical Actuation of Movement (n=11)	Autonomous Movement (n=3)
Transfer (n=25)	3.83	3.00	2.67	2.00
Translate (n=18)	1.07	0.50	2.67	0.33
Rotate (n=18)	1.07	1.00	1.20	0.50

Values can be interpreted as the odds of a code being applied when each function was used. There was not a consistent trend for these Function-Code pairs. For research question 2, seeing "Sensor" applied when sense was used was 3x as likely as when sense was not used. It should be noted that "Sensor" was applied 15 times and all 15 times the sense function was used. The 3x estimate is a product of the Haldane correction. Sense was used a total of 27 times.

For research question 3, seeing “Supportive Object” applied when support was used was 2.5x *less* likely as when support was not used. “Supportive Object” was only applied seven times, and five of those times support was not used. Support was used a total of 18 times.

For research question 4, seeing “Computer” applied when process was used was 2.5x as likely as when process was not used. “Computer” was applied 7 times, and five of those times process was used. Process was used a total of 21 times.

Research question 5 is addressed in Table 49.

Table 49: Co-occurrence odds for control-related functions with “Physical Actuation of Movement.”

	Physical Actuation of Movement (n=11)
Actuate (n=25)	10.00
Regulate (n=16)	1.20
Change (n=14)	1.75
Stop (n=14)	0.57

Actuate was strongly associated with physical actuation of movement. For research question 6, seeing “Provides Visual/Auditory/Tactile Feedback” when indicate was used was equally as likely when indicate was not used. This code was only applied twice.

6.5.2 Model Population Summary

Table 50 shows the distribution of group membership for the model population. Each cell corresponds to the number of individuals assigned to each group during clustering. Cognitive tasks refer to membership for both “Applying Existing Knowledge” and “Problem-solving and Decision-making.”

Table 50: Model population group membership by task.

Group #	Gross Upper-body Movement	Fine Motor Movement	Visual Discrimination	Auditory Discrimination	Cognitive Tasks
1	264	471	518	336	270
2	217	145	98	226	246
3	135	-	-	53	99

Group membership was predominately distributed to the lower risk groups. Spearman’s rank order correlation coefficient was used to examine correlation between group membership (Table 51). A value of 1 indicates perfect positive correlation between low and high-risk group membership. All cluster memberships demonstrated a significant, but weak – moderate association.

Table 51: Spearman's rank order correlation coefficient (p-value) for group membership.

	Gross Upper-body Movement	Fine Motor Movement	Visual Discrimination	Auditory Discrimination
Fine Motor Movement	0.425 (<.001)	1		
Visual Discrimination	0.284 (<.001)	0.104 (.009)	1	
Auditory Discrimination	0.140 (<.001)	0.246 (<.001)	0.158 (<.001)	1
Cognitive Tasks	0.202 (<.001)	0.213 (<.001)	0.198 (<.001)	0.189 (<.001)

6.5.3 Optimization Results

The optimization model was first run with $\alpha = 0.5$. HD and SS varied insignificantly across population sizes. NS, however, increased and plateaued over sizes. The smallest population size with the largest NS in the plateau region was selected. Figure 15 displays the pareto front alongside student solutions.

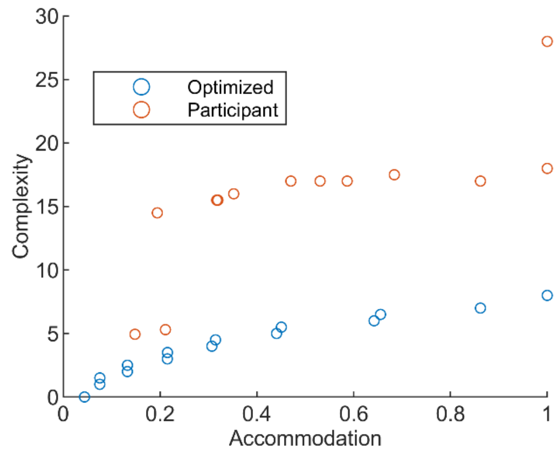


Figure 15: Pareto front for alpha = 0.5 plotted against student responses.

The optimization model produced solutions with better trade-offs in all cases. Table 52 contains a summary of coded participant responses to the question "What logic did you use when selecting human or machine assignments for each function?" Product cost and user capability was the most citied justification for allocation decision-making.

Table 52: Coded responses for manual function allocation optimization exercise question, n (%).

Production Cost	**User Capability**	**Feasibility of Design**	**Convenience to User**	**Reliability of System**
7 (44%)	7 (44%)	4 (25%)	3 (19%)	3 (19%)

Sensitivity Analysis: To evaluate the influence of alpha on results, the model was run with alpha varying between 0 and 1. NS, HD, and SS were used as evaluation criteria. If NS did not coverage at 8,000 population size, larger sizes were evaluated in increments of 500. Table 53 summarizes the output at different alpha levels. As alpha increased, the number of viable solutions dramatically increased.

Table 53: Sensitivity analysis results for alpha parameter.

Alpha	**Pop Size**	**NS**	**HD**	**SS**	**Product Family Member Quantity Distribution**				
					1	**2**	**3**	**4**	**5**
0.1	4500	21	0.339	0.032	3	17	1	0	0
0.25	5500	21	0.340	0.030	4	15	2	0	0
0.5	7000	22	0.343	0.029	5	16	1	0	0
0.75	7500	62	0.344	0.017	7	46	9	0	0
0.9	7500	104	0.35	0.014	2	80	19	3	0

NS = # unique solutions, HD = hyper-area difference, SS = solution spread

Figure 16 displays the pareto fronts for alpha values of 0.1 and 0.9. There was as significant amount of overlap between these two pareto sets. Both alpha levels seemed to form discontinuous, micro-pareto fronts.

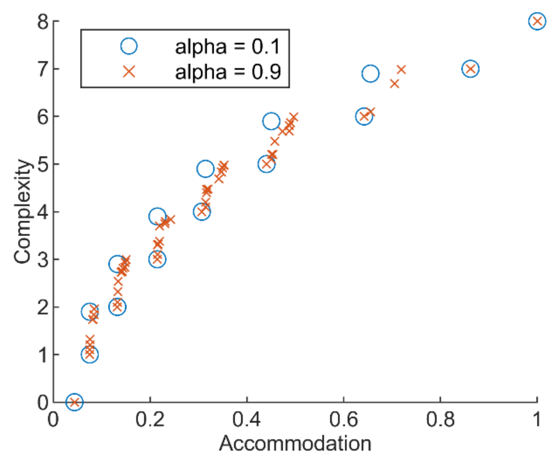


Figure 16: Pareto fronts for alpha = 0.1 and alpha = 0.9.

6.5.4 Optimization Model Sample Solutions

Two output solutions corresponding to high complexity, high accommodation and low complexity, low accommodation are shown. Solutions from alpha = 0.9 were used because it provided a large amount of solution variety.

<u>Example #1</u>: Shown in Table 54 are the allocations for a solution that provided 0.705 accommodation at 6.69 complexity. Figure 17 displays average user

performance values for tasks by product family member. Tasks are abbreviated as follows: GUBM = Gross upper-body movement, FMM = Fine motor movement, VIS = Visual discrimination, AUD = Auditory discrimination, KNW = Applying Existing Knowledge, and DM = Problem-solving and decision-making. This solution set produced two product family members. Product #1 included additional sensing functions, while product #2 included machine storage of information.

Table 54: Example solution #1 human (H)-machine (M) allocations.

	Product Family Member (User Count)	
Function	1 (n = 337)	2 (n = 279)
Sense Signal	**M**	H
Process Signal	**M**	**M**
Store Signal	H	**M**
Extract Material	**M**	**M**
Transfer Material	**M**	**M**
Store Material	**M**	**M**
Detect Signal	**M**	H
Process Signal	**M**	**M**

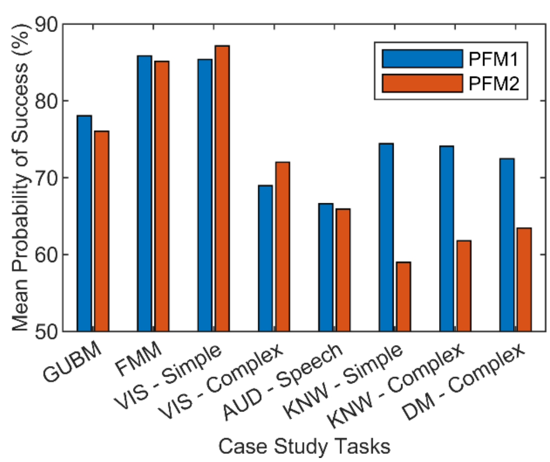


Figure 17: Population mean task performance for example #1. PFM = product family member.

Example #2: Shown in Table 55 are the allocations for a solution that provided 0.317 accommodation at 4.47 complexity. Figure 18 displays average user performance values for tasks by product family member. This solution produced four product family members, with the largest differences being between functions associated with sensory and cognitive tasks.

Table 55: Example solution #2 human (H)-machine (M) allocations.

	Product Family Member (User Count)			
Function	1 (n=253)	2 (n=243)	3 (n=5)	4 (n=115)
Sense Signal	H	H	H	**M**
Process Signal	H	H	H	H
Store Signal	H	**M**	**M**	**M**
Extract Material	**M**	**M**	H	H
Transfer Material	**M**	**M**	**M**	**M**
Store Material	**M**	**M**	**M**	**M**
Detect Signal	H	H	**M**	H
Process Signal	**M**	H	H	**M**

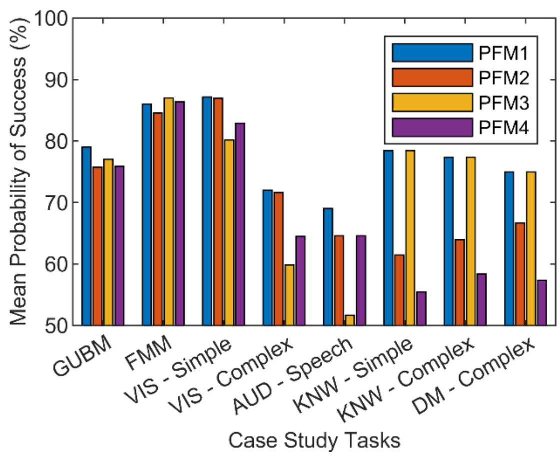


Figure 18: Population mean task performance for example #2. PFM = product family member.

6.6 Discussion

The results of the modeling and optimization approach are discussed in this section.

6.6.1 Human-Machine Function Modeling Validation Outcomes

The results of first two participant exercises revealed few differences in outcomes between the conventional function modeling perspective and the human-machine modeling perspective, and none that demonstrated significant differences. Ultimately, these exercises did not demonstrate what was hypothesized, that the proposed approach would promote accessible design thinking in concept generation. There were several issues that were encountered that may have contributed to this. This study was in the final stages of development when the COVID-19 virus first surged in the United States, and universities were forced to transition student activities to online. This presented the following challenges: 1) The study was forced to transition from laboratory to online format; 2) Recruitment was negatively impacted; and 3) Challenges related to participant anonymity and effort were encountered.

Online Format: The first issue was forcing the study to transition to an online format when it was originally intended to be performed in a laboratory setting. This necessitated significant simplification of modeling procedures to conform with the capabilities of the online survey platform used (Qualtrics). Further, participants were unable to seek clarification on the task if they encountered difficulties, as would be possible in a supervised setting.

University Shutdown and Recruitment: Next, it is believed that the university shutdown had a significant impact on recruitment. Ideally, each participant would only perform the conventional approach or the proposed approach so that outcomes could be observed independently. However, when it was determined that recruitment was likely to be sparse, it was decided that participants should perform each exercise in sequence to ensure both were represented. This created an anchoring bias, where participant responses deviated very little in the proposed approach compared to those in the conventional approach. The order of the exercises could have been randomized, however it was unknown at the time whether ordering effects could be detected given anticipated recruitment.

Participant Anonymity: Finally, it is believed that the anonymous (to course instructors) and online nature of the exercises may have led to complacency in responses by participants due to lack of accountability (often referred to as Insufficient Effort Responding - IER). This is consistent with past research (Camus, 2015; Meade & Craig, 2012). Many responses by participants demonstrated lacking effort. In some cases, responses were nonsensical or repeated verbatim. Several participants data (n=2) were removed from the analysis because the quality of responses made them unusable. It is suspected that if the study were performed in a lab setting as intended then the responses would have been of higher quality and effort.

Lessons Learned: There were several lessons learned from this work regarding implementation of design studies in online formats. First - when anonymous participants have a choice in a design exercise, they will likely make the choice that requires the least effort (e.g., # of product family members). Choices of these nature must be limited. Next – when implementing a design study online, the limitations of the participant population should be considered. Undergraduate students (typically) only have experience applying design methodology in academic settings. This means that activity requirements must be specific and not overly complicated. This study could have been modified by removing any feature that did not directly contribute to testing the core hypothesis. For example – the "information processing functions" may have been an unnecessary feature. The purpose of these mandatory functions was to supplement engineering student understanding of human cognition, however it may have contributed to overloading students with information. Further, this population has concentrated priorities on academic achievement. It should be considered whether this priority can override the desire to provide thoughtful responses, and instead encourage responses that maximize the efficiency of incentives by minimizing time allocated to the task. One option could be to increase the amount of extra credit given. 0.5% on the final grade would create a more tangible impact for students. Another option could be to make the assignment a required homework. Integrating the exercise into the course could motivate students to increase effort.

6.6.2 Population Model

Using NHANES data to model the population and expert elicitation to quantify population task performance was a useful exercise to understand the heterogeneous capabilities of the user population. Each task-specific user group was well represented in the population (Table 50), indicating that this population is indeed heterogeneous. Further, weak-moderate correlation between risk-levels was observed (Table 51), indicating that individuals who perform poorly on one type of task may be more likely to perform poorly on another. The strongest association was observed between gross upper-body movements and fine motor movements. This is not surprising as these tasks shared several of the same clustering variables and require overlapping motor skills. These associations create a complicated landscape of potential design requirements that is difficult to comprehend. Thus, an automated decision-aid for configuring human-task requirements, such as an optimization model, seems justified.

6.6.3 Optimization Model

The optimization model successfully produced a variety of potential function allocation solutions. Figure 15 shows the pareto front for alpha = 0.5 (i.e., MFC and UFC are equal). The front is well distributed in the objective space, with candidate solutions spanning the max and min value for each metric. The shape of the pareto front contained discontinuous groups of solutions. This can be attributed to the discrete performance values assigned to population individuals. Individual task performance was assigned based on the 50th percentile for groups across all individuals, resulting in a discrete set of potential total accommodation values. Task performance could have also been modeled as a random variable, where individual performance was drawn from a beta distribution based on the expert provided confidence intervals. It is suspected that this would produce a more continuous pareto front. Solutions in close, local proximity do not necessarily represent similar design solutions. For example, the two solutions at ~0.6 accommodation in Figure 16 are quite different, with one including three product family members and the other only including one.

<u>Participant Performance</u>: The optimization model produced better results than participants in all cases. Participants produced results ranging in accommodation but were unable to do so with the same efficiency as the model, despite indicating that

cost and user capability were top priorities when allocating functions (Table 52). This demonstrated that, without the tools demonstrated in the paper, individuals manually allocating functions are unlikely to be efficient with respect to accommodation and product complexity. The only materials participants were provided with was summary statistics on the population of interest. The model had information that was not provided to participants. This information would not be easily accessible or digested without computational tools. The participants did not have information on function-task mappings, population task performance values, and exact metrics for evaluation, which was input to the optimization model. It is possible that with this information participants may have performed differently, however this would not reflect the typical scenario in practice. Collectively processing these model inputs and determining how they relate to function allocation is a complicated task that seems to warrant algorithmic solutions.

Sensitivity Analysis: The sensitivity analysis demonstrated that the weighting of complexity metrics can have a significant influence on model output (Table 53). Across all alphas, HD and SS changed only marginally. As alpha increased past 0.5 (MFC > UFC), however, the number of viable solutions increased dramatically. It also encouraged larger product families. When MFC (the "cost of each unit") becomes more important, the efficiency of tailoring additional products to individuals increases. When alpha was decreased below 0.5, the number of unique solutions did not change significantly. Comparing alpha = 0.1 (Figure 16) to alpha = 0.5 (Figure 15), the spread between solution pairs is different. It seems that as alpha decreases, this spread increases, indicating that small, "local" increases in accommodation became more costly. This is likely due to this increased cost of introducing a previously unused function.

Example Solutions: The example solutions presented demonstrate that a variety of solutions are possible, and that product families can be leveraged to produce efficient allocations. The product family strategy metric successfully encouraged the use of platforming in solutions. In example #1, a platform emerged spanning five of the eight functions, with only minor alterations for individual products. In example #2, a smaller platform emerged across all products, containing only two functions.

6.6.4 Proposed Use

The intended use of the methodology is for early-design decision-making for function allocation. Intentional or otherwise, required human functioning is a design decision. This methodology can be reproduced to navigate the trade-offs between accessibility and cost resulting from these decisions. The output of the optimization model produces varied solutions that can be independently evaluated and selected based on stakeholder values. This solution can be used to drive concept development. This approach could be used for nearly any system, however it is most useful for highly heterogeneous populations where manually evaluating population task performance becomes complicated because of the wide array of potential use cases. Also, this approach is especially important for safety-critical systems, where human functional failure could result in injury or death.

This modeling approach should be used during concept development, prior to any physical product design. Further, it should be used as a precursor to detailed functional modeling. As discussed before, not all human functions have a machine function that serves as a 1-to-1 replacement (Dekker & Woods, 2002). In practice, a product may require supportive or auxiliary functionality. When a human performs a task, they have the entire body at their disposal, which implicitly contains some of these supportive functions (e.g., the body generates, stores, and supplies energy). Designers should use the output of this modeling approach as baseline to build on.

Chapter 7: Conclusions

7.1 Summary of Contributions

7.1.1 Leveraging Physician Expertise and National Population Data to Model Heterogeneous Population User Groups

In this chapter, a combined expert and data-driven approach for modeling product user groups was proposed. The approach was specifically developed to guide product personalization. A taxonomy of tasks for describing product interaction was introduced, and the taxonomy was translated into tasks using language to facilitate judgment from physicians regarding task performance (P&C Physician Judgment Tasks). These tasks then guided identification of performance-driving characteristics and acquisition of data for user group clustering. The approach was demonstrated on the diabetes population case study, where task-specific user groups were identified for six generic tasks required for medical device interaction. Data was retrieved from the NHANES database, guided by input from internal medicine physicians. The output of this approach was task-specific patient user groups that can be used to guide the customization of medical devices. Understanding the needs of product users in safety-critical domains and incorporating these needs into the design process is critical for safe and effective products. This work provides designers a novel and cost-effective means to characterize user sub-populations as a basis for targeted personalization.

7.1.2 Quantifying Human Performance for Heterogeneous Population User Groups using Expert Elicitation

Studying heterogeneous user populations for product design is difficult because of the wide array of use cases that may present themselves, thus placing a high bar on recruitment and experimental efforts. In this work, a general process to quantify human performance in heterogeneous user populations was proposed. The process suggests that expert elicitation can reduce the burden of quantifying performance by reducing the need for recruiting users. This approach was demonstrated on the diabetes population case study, focusing on tasks required for medical device interaction. Results demonstrated that experts could discriminate user performance across task risk-levels, and for similar tasks under different conditions. The needs of vulnerable users in heterogeneous populations have gone inadequately addressed in the past, with system designers relying on 1-size-fits-all approaches or minimally differentiated products. This work demonstrated a cost-effective approach to quantify human performance and risk that can be used to guide safe and accessible design.

7.1.3 Optimizing Function Allocation for Accommodation of Heterogeneous Populations

A function allocation optimization model for early design stages was proposed. This model relied on an adaptation of traditional functional modeling approaches, where functions could be allocated to humans or machines. Two metrics to evaluate allocations were introduced (accommodation, complexity) and formulated into a multi-objective optimization model. The model was demonstrated on the diabetes population case study, where functioning for a diabetes self-management device was optimally allocated. Student participants were recruited to perform several exercises. The first two exercises served to validate the function modeling framework, and the third exercise sought to validate the optimization model. The function modeling validation exercises did not effectively demonstrate differences in performance between the conventional and proposed modeling approaches. This was primarily attributed to switching from an in-person study to an online study, as necessitated by the COVID-19 virus outbreak. While this part of the study did not demonstrate the anticipated effects, several lessons-learned regarding online study design for engineering design studies were obtained and discussed. For the third exercise, participants were unable to perform better than the optimization model in all cases, demonstrating the utility of the optimization model compared to uninformed allocation of functions. This approach can be replicated for virtually any system but is particularly suited for systems with heterogeneous user populations. The output of the model can serve as a baseline for more detailed function modeling, or to facilitate general concept ideation. Evaluating the capabilities of the intended user population in early design stages is critical to mitigate costly redesign given new information in later stages, and critical to minimize error by end users. The proposed modeling approach facilitates these considerations.

7.2 Overarching Implications

The case study data produced in this dissertation can be applied to patient-facing medical devices used by diabetes patients, but the methodology can be replicated for user populations and products in other domains. For all cases, internal medicine physicians should still be relied on as experts capable of making judgments regardless of the domain.

The ease of replicating this approach for other populations primarily depends on the data available in NHANES. For other chronic disease populations, this process can easily be replicated by filtering NHANES participants who reported the disease. For example, prior work demonstrated the clustering procedure on the hypertensive population (Knisely & Vaughn-Cooke, 2020a). For the general consumer population, all NHANES participants can be used. For occupation-specific products with entry requirements (e.g., provider-facing medical devices, mining equipment, aviation equipment), NHANES data can be filtered by occupation. Note, however, that not all years of NHANES include detailed survey questions on occupation. Additionally, for occupations with few workers relative to the general population, participants may be too sparse in the dataset to allow clustering. For example, NHANES 2013-2014 contains 519 participants who reported working in healthcare, however only 18 who reported working in mining. For cases such as the latter, the general population can be assumed, or creative sampling based on industry demographics can be used as demographic data is plentiful in NHANES. Otherwise, alternative means to define task-specific user groups should be identified. The other elements of the methodology can be replicated regardless of the product domain.

Designers should use this methodology in the conceptual phase of design to establish the baseline differentiation of product functionality for products in a product family. This could be thought of as the first “layer” of differentiation, where each functional variant can go on to drive additional differentiation for each user group identified. The first step should be to take the output product functions from this methodology and build out a more detailed function structure. Next, these function structures can guide product embodiment, where physical design solutions are generated to fulfill system functions. To establish a product platform, design solutions may be shared across product family members that share functionality. As design embodiment becomes more detailed, and product design parameters are established, additional differentiation can occur based on scaling parameters to meet sub-user group population needs. Figure 19 demonstrates this progression.

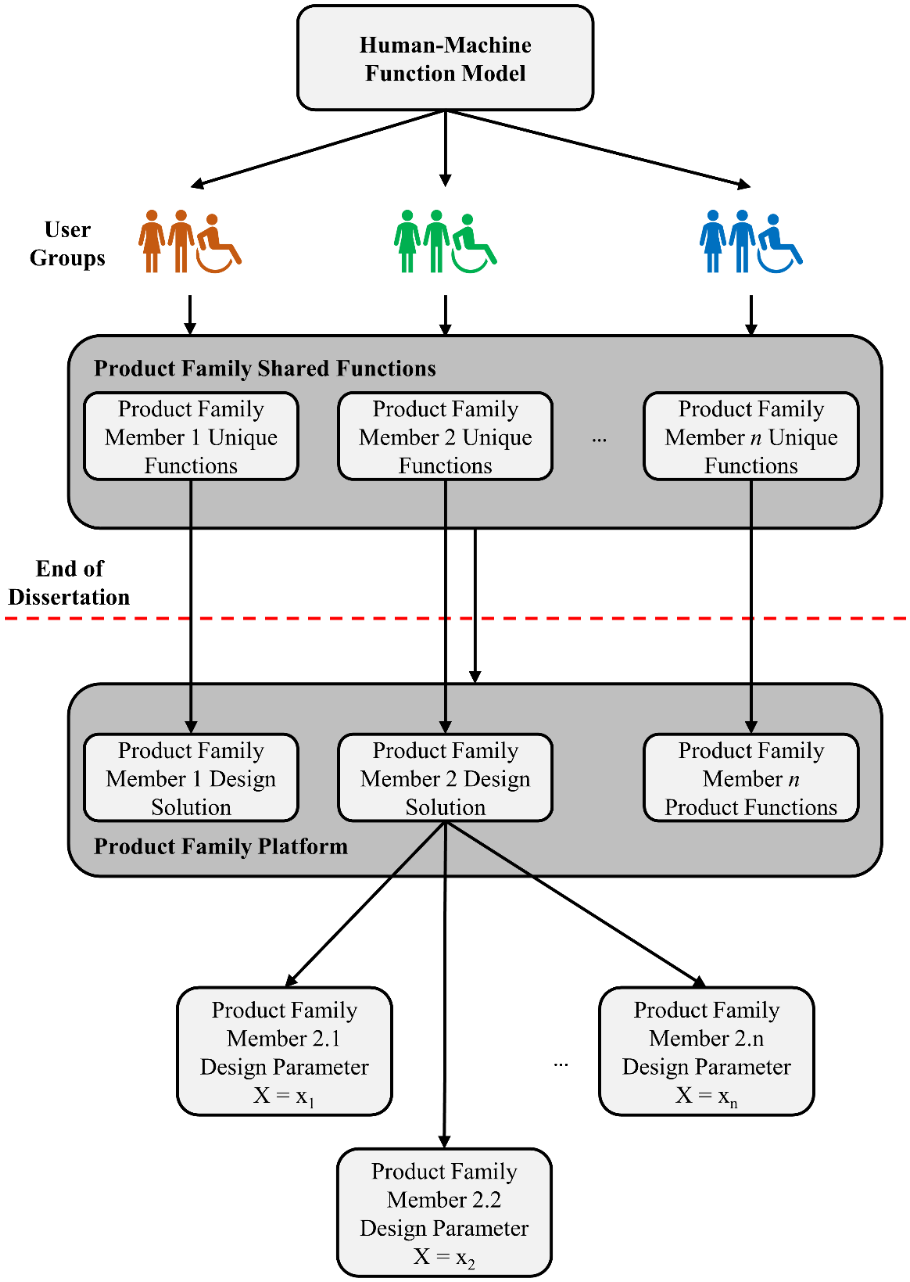


Figure 19: Proposed use of methodology output.

The case study data produced from this dissertation also has utility. The output of each chapter can be used by engineers and designers targeting the diabetes population. The user groups and performance values can be used for any product where the core users are individuals with diabetes. This would primarily include other medical devices, such as insulin pumps or health information technology systems. Designers would need to identify the core tasks relevant for the given product and use this to identify relevant task-specific user groups. Designers could then follow the modeling process detailed in Chapter 6 to create a function model for their case and replicate the optimization procedures to output candidate function allocations. A more detailed discussion on the proposed use of each chapter's output can be seen in the Proposed Use sections of Chapters 4-6.

7.3 Methodological Validity

This section provides a brief discussion of the efforts taken to validate this methodology as well as areas that lack validation and should be subject to future work. Total method validation is important to address for several reasons. First, validation can provide confidence that the method is beneficial with respect to the established objectives, and that the observed benefits can be repeatedly obtained in practice. Second, *total* method validation must be specifically addressed, as validation of individual method elements does not guarantee validity of the entire process.

Validity will be discussed in the context of the Validation Square, a framework proposed to facilitate structured validation of design methodology (Seepersad et al., 2006). The Validation Square seeks to address the challenges associated with validating design methodology in a research context, primarily being the difficulty of following a proposed design solution through the complete product life cycle. The Validation Square provides a process to facilitate structural validation (method provides solutions correctly) and performance validation (method provides the correct solutions). The process suggests the following tasks to ensure method validity:

1. **Domain-Independent Structural Validity** – The method is determined to be logical, internally consistent, and mathematically correct. This includes the internal consistency of "parent" constructs that are considered influential to the method, as well as the internal consistency of the method itself.
2. **Domain-Specific Structural Validity –** An example problem is identified and justified as appropriate given the context of the method.
3. **Domain-Specific Performance Validity –** The method provides useful results with respect to the stated purpose. This includes applying the method to the example problem, defining and applying metrics of usefulness to the example problem, and demonstrating that the usefulness is directly a result of applying the method.
4. **Domain-Independent Performance Validity –** The method is reasoned to be useful beyond the example problem.

The proposed method will now be discussed in the context of each of these steps, identifying where each step was addressed and highlighting elements that are missing. The primary focus will be on the validity of the methodology as a whole. The validity of the individual constructs will be discussed as well, broadly corresponding to modeling user groups (Chapter 4), quantifying user group task performance (Chapter 5), the human-machine functional modeling approach (Chapter 6), and function allocation optimization (Chapter 6).

Domain-Independent Structural Validation – The first step for this element of validation is to identify the requirements and need for the design method. Chapter 1 of this dissertation provided details on the purpose and requirements for the overall method. Further, for each primary construct that make up the method, a review of the relevant "parent" constructs was provided in Chapter 2. This includes highlighting the limitations of these constructs, therefore justifying the need for each individual proposed construct. Next, the internal consistency of the method is established in Chapter 1.2. Each step of the method is explained in the context of the greater method goals, including sub-steps, inputs, and outputs within and between constructs. Further, a flowchart is used to demonstrate the logical flow of each method construct (Figure 1).

Domain-Specific Structural Validation – Chapter 3 of this dissertation introduces a case study (example problem) along with the justification of its appropriateness for the domain. As this method revolves around addressing the needs of specific user populations, the appropriateness of the example problem depends mostly on the characteristics of the target population. The inclusion criteria for example problems are listed at the beginning of Chapter 3.

Domain-Specific Performance Validation – The method was applied to the diabetes self-management device case study proposed in Chapter 3 to demonstrate its performance. The Validation Square process suggests identifying metrics for usefulness to measure if the method satisfies its intended purpose, as well as evaluating if the demonstrated usefulness is linked to applying the method. This is discussed for each of the primary method constructs.

Modeling User Groups – To validate the clusters produced by this process, four quantitative metrics (silhouette index, CH index, connectivity, BIC) were used to evaluate cluster quality. Additionally, clusters were subjectively evaluated for qualitative separation based on the researchers prior understanding of user risk. The usefulness of these clusters relies on the assumption that the expert-driven process for selecting meaningful NHANES variables was successful, and that the diabetes patients sampled by NHANES reflect the true diabetes population. While we can be relatively confident about the latter, future work should seek to address the validity of the former.

Quantifying User Group Task Performance – The validity of the expert elicitation relies on the previous efforts to validate the elicitation protocol used (Cooke protocol) and the quantitative metrics used for expert evaluation (calibration score, information score) (Colson & Cooke, 2017). Acknowledging the limitations discussed in Chapter 5.5, this provides confidence that the elicited values should be better than a layman estimate, though future efforts should be taken to validate the elicited values in the true population, discussed further in the 7.4 Future Work. The result of the elicitation also provides additional confidence that the subjective evaluation of cluster separation was meaningful, as expert elicited values for user groups corresponded with the stated qualitative risk-levels by the researchers.

Human-Machine Function Modeling – The proposed modeling approach was based off prior modeling approaches and adjusted given the objective of function allocation. The purpose of the modeling approach was to facilitate function allocation in the early-stage product design context, and to improve student concept generation with respect to product accessibility. While efforts were taken to validate this hypothesis (student exercises 1&2, Chapter 6), these efforts did not ultimately produce different outcomes. As discussed in Chapter 6, we do not believe these outcomes are representative due to several limitations resulting from the COVID-19 pandemic and University shutdown. Future work should seek to perform this study again under better conditions, discussed further in Chapter 7.4.

Function Allocation Optimization – The quality of the optimization model output was evaluated using several quantitative metrics (hyper-area difference, solution spread, unique solutions). These metrics only measure the quality of the multi-objective output based on theory surrounding pareto optimality with respect to the defined objective functions. This does not necessarily mean produced solutions will be useful in real-world applications. The usefulness of solutions relies on the validity of the objective functions used, which were generated uniquely for this dissertation. The validity of the accommodation metric primarily relies on the validity of the methodological steps taken prior (performance data, function model produced), while the complexity metric is solely based on prior literature (Gill et al., 2017) and logical reasoning given the properties of product families. Future validating work is required to develop confidence in the usefulness of these complexity metrics. For example, these metrics could be applied to real-world products, and the resulting values could be compared to the cost of the product.

Further, the necessity of the optimization model in isolation should be investigated. In the third student design exercise, students completed the function allocation task with only information that would be available to them in typical engineering practice. Therefore, they did not have access to the information produced by this method (user groups, user group task performance estimates) that was used as input to the optimization model. This allowed us to compare participant performance against the method in entirety, but not to the model in isolation. To determine if the optimization model is necessary for selecting optimal design solutions in isolation, participants would need to complete the same task given the same information that was input to the model, including data produced in the prior sections and as well as the nature of the objective functions. This would validate that a computational solution was warranted for generating and evaluating design solutions with respect to the objective functions.

Complete Method - The method in its entirely was quantitatively validated during the third student design exercise where students were asked to perform a manual function allocation task. This exercise demonstrated that the data and tools produced in this dissertation can produce outcomes that are more efficient with respect to the defined objectives than student engineers. Once again, this relies on the validity of these objective metrics, which must be tested in practice. This does not guarantee that the functional requirements produced will actually produce a product that is highly usable and cost-effective. However, based on the observed results and the discussed validity, it can be reasoned that these efforts would produce better outcomes than if the methods had not been used at all for this design case study.

<u>Domain-Independent Performance Validation</u> – Given the previous discussions, generalizability of observed outcomes is assumed for design problems with similar characteristics. The chosen design problem was deliberately selected

given that it exhibits the properties discussed in Chapter 3 (heterogeneous population, safety-critical domain). Steps were taken to ensure the method was applied to the design case study in an internally consistent manner, with efforts taken to validate the usefulness of the method in practice. It is therefore reasonable that, while not explicitly tested, the observed benefits are expected to transfer to design problems with similar properties.

7.4 Future Work

There are several opportunities for future work that could stem from this dissertation. In all engineering design problems, there are degrees of uncertainty that can influence expected system performance in practice and should be accounted for (Cuneo et al., 2017; Kota & Chakrabarti, 2010). In design methods, it is important to quantify the uncertainty associated with each methodological stage because this uncertainty can propagate throughout the method, which can lead to overconfidence in the intermediate and final results. This design methodology consists of several stages that are linked via inputs/outputs. Future work should seek to address uncertainty associated with method outputs that have not already been addressed and should seek to formally evaluate how that uncertainty propagates to the output of the optimization model. Of particularly note is the uncertainty estimates obtained in the expert elicitation. It should be investigated how elicited uncertainty can be integrated into a robust version of the multi-objective optimization problem (He et al., 2019).

Further, while internal validation was well covered, there was a lack of external validation for the proposed methodology, as discussed in the prior section.

Efforts should be taken to validate the performance values generated for each task-specific user group. Participants could be recruited using the user group characteristics and proportions identified as benchmarks. These benchmarks can be monitored during the recruitment process and recruitment strategies can be updated to target segments that are not well represented. Targeting users based on these characteristics would be more challenging than typical efforts that rely on demographics and could be an area of future research itself.

Recruited participants could then perform controlled laboratory versions of each task. Performance quantities could be experimentally measured and compared to the values output during the expert elicitation. If values were close and correlated, this would provide confidence that physicians were able to make reasonable estimations for task performance, and that the process in Chapter 4 produced characteristics that are reasonable predictors of task performance.

Efforts should also be taken to re-validate the human-machine function modeling procedure developed for Chapter 6. Due to the limitations discussed previously, it is currently unclear if the proposed function modeling procedure was useful for promoting accessible design decision-making. Conducting a similar study in a laboratory setting could overcome the limitations encountered with the remote study design.

Other work could expand on and demonstrate how the output of the optimization model could be used. An accompanying methodology that takes the output functions as input and translates them into a physically realized product family (akin to Figure 19) would help turn the proposed approaches into a complete design methodology. Testing these products in the intended user population and measuring performance could be used to validate that the utility of design-decisions made early in the design process persist into later design stages. Further, it would provide evidence that these early design tools have practical utility in contributing to a complete design methodology.

Appendices

Appendix A: Patient Characteristic – NHANES Variable Mapping

Table 56: User characteristic **-** *NHANES variable mapping for "fine motor movement" (adapted from (Knisely & Vaughn-Cooke, 2020a)).*

Expert Survey Characteristic	**NHANES Variable**	**Format**	**Justification**
Age	Age	Continuous	Age and decreased hand mobility are associated (Seidler et al., 2010).
Physical Independence	Reported difficulty using fork, knife, or cup	Ordinal – No difficulty, Some difficulty, Much difficulty, Unable to do, Does not do	Activity is a specific case of "fine hand manipulation"
Physical Independence	Reported difficulty grasping/holding small objects	Ordinal – No difficulty, Some difficulty, Much difficulty, Unable to do, Does not do	Activity is a specific case of "fine hand manipulation"
Disease History	Reported having: - Arthritis - Gout - Bone/joint injury - Neck and Back Problem	Binary – Yes or No	"Find hand use" linked with post-acute musculoskeletal disease ICF core set (Scheuringer et al., 2005).
Disease History	Reported having: - Congestive heart failure - Angina/angina pectoris	Binary – Yes or No	"Fine hand use" linked with cardiopulmonary post-acute ICF core set (Wildner et al., 2005).
Disease History	Reported having a stroke	Binary – Yes or No	"Fine hand use" linked with stroke ICF core set (Geyh et al., 2004).
Exercise	Reported physical activity at work	Ordinal – None, Moderate, Vigorous	Physical activity is associated with fine motor skill (Liubicich et al., 2012; Miyake et al., 2013).
Exercise	Reported physical activity recreationally	Ordinal – None, Moderate, Vigorous	Physical activity is associated with fine motor skill (Liubicich et al., 2012; Miyake et al., 2013).

Table 57: User characteristic - NHANES variable mapping for "visual discrimination."

Expert Survey Characteristic	NHANES Variable	Format	Justification
Age	Age	Continuous	Age and decreased vision are associated (Rudman et al., 2016).
Disease History	Reported difficulty seeing with or without correction	Binary – Yes or No	Activity describes the task.

Table 58: User characteristic - NHANES variable mapping for "auditory discrimination."

Expert Survey Characteristic	NHANES Variable	Format	Justification
Age	Age	Continuous	Age and decreased hearing are associated (P. Wu et al., 2020).
Disease History/Severity	Reported hearing quality without correction	Binary – Yes or No	Activity describes the task.
Disease History	Wears a hearing device (aid, amplifier, or implant)	Binary – Yes or No	Included to compliment prior characteristic.

Table 59: User characteristic - NHANES variable mapping for "applying existing knowledge" and "problem-solving and decision-making."

Expert Survey Characteristic	NHANES Variable	Format	Justification
Age	Age	Continuous	Age is associated with cognitive decline (Meng et al., 2017).
Memory, Attention, and Decision-making Skills	Problems managing money	Ordinal – No difficulty, Some difficulty, Much difficulty, Unable to do, Does not do	Activity is a specific case of “applying existing knowledge” and “problem-solving and decision-making”.
Memory, Attention, and Decision-making Skills	Reports experiencing confusion/memory problems	Binary – Yes or No	Memory is an integral part of “apply existing knowledge” and is associated with “problem-solving and decision-making” (Del Missier et al., 2015).
Memory, Attention, and Decision-making Skills	Reports having serious difficulty concentrating, remembering, or making decisions	Binary – Yes or No	Variable describes the task.

Table 59 (continued)

Memory, Attention, and Decision-making Skills	Reported feeling tired or having low energy over the last two weeks	Ordinal – Not at all, Several days, More than half, Every day	Low energy is associated with poor memory and decision making (McCoy & Strecker, 2011; Whitney et al., 2015).
Disease History	Reported having a stroke	Binary – Yes or No	"Focusing attention" and "Solving problems" linked with stroke ICF core set (Geyh et al., 2004).
Psychiatric Disorder	Reported frequency of feeling worried or anxious	Ordinal – Never, Few times a year, Monthly, Weekly, Daily	"Solving problems", "decision-making", "attention function", and "memory function" all linked with mental disorder ICF core set (Guilera et al., 2020).
Psychiatric Disorder	Reported severity of anxiety	Ordinal – A little (mild), A lot (severe), or somewhere in-between (moderate)	"Solving problems", "decision-making", "attention function", and "memory function" all linked with mental disorder ICF core set (Guilera et al., 2020).
Psychiatric Disorder	Reported frequency of feeling depressed	Ordinal – Never, Few times a year, Monthly, Weekly, Daily	"Solving problems", "decision-making", "attention function", and "memory function" all linked with mental disorder ICF core set (Guilera et al., 2020).
Psychiatric Disorder	Reported severity of depression	Ordinal – A little (mild), A lot (severe), or somewhere in-between (moderate)	"Solving problems", "decision-making", "attention function", and "memory function" all linked with mental disorder ICF core set (Guilera et al., 2020).

Appendix B: GMM Model Performance Comparison

Cluster BIC values for “gross-upper body movement”, “fine motor movement”, and cognitive tasks are shown in Figures 20-22. For some model and cluster count combinations, there was issues with convergence. Results for these cases are not shown. For more detail on models (EEI, EII, EVI, VEI, VII, VVI), see (McParland & Gormley, 2015). Note that in this package, BIC is formulated such that the maximum value is sought, while in many other cases it is formulated such that the minimum is sought.

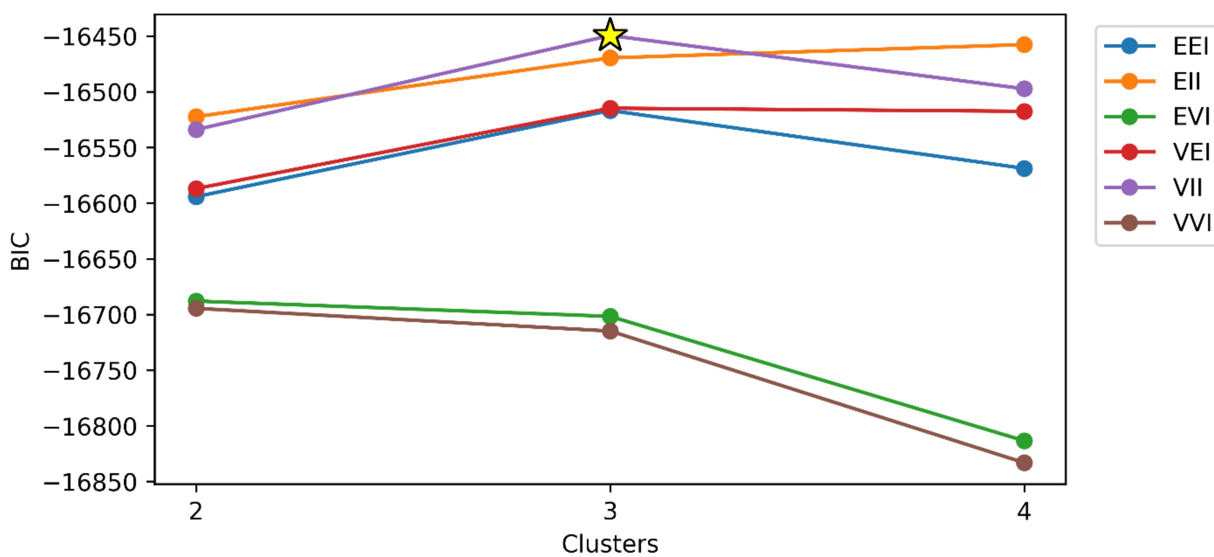


Figure 20: GMM BIC values for gross-upper body movement clusters. Star indicates highest value.

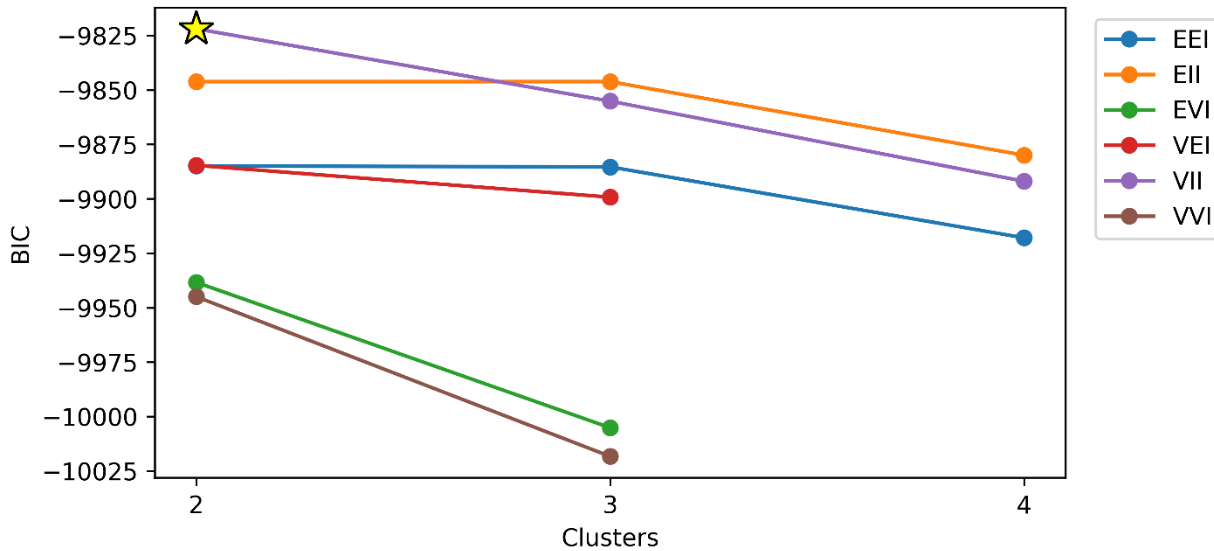


Figure 21: GMM BIC values for fine motor movement clusters. Star indicates highest value.

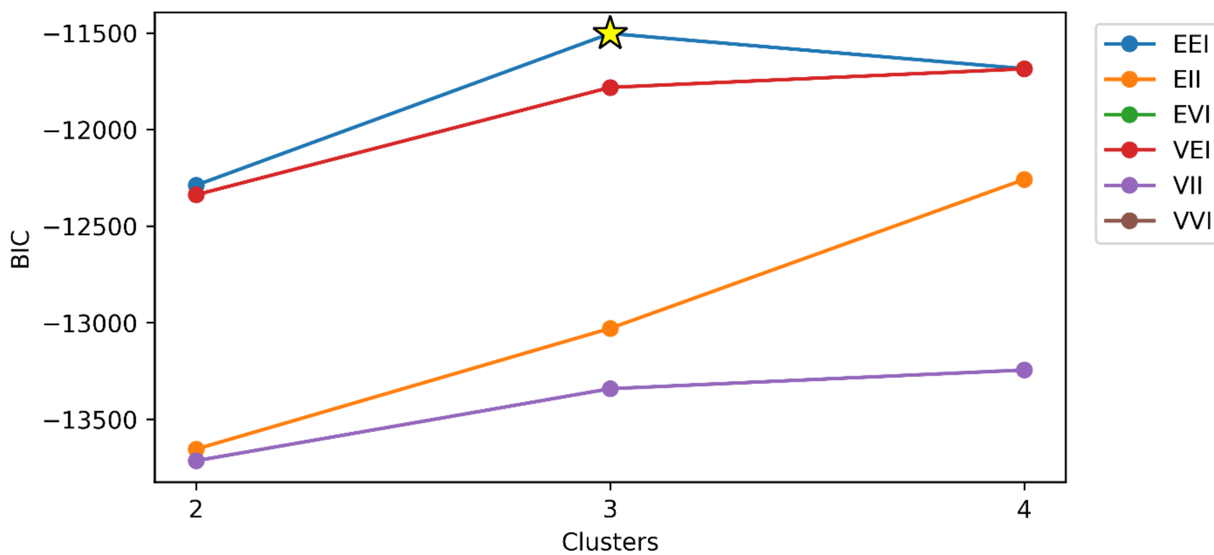


Figure 22: GMM BIC values for cognitive tasks' clusters. Star indicates highest value.

Appendix C: Cluster Dominant Characteristic Summary

Tables 60-62 display summaries of dominant cluster characteristics for each cluster solution. Characteristics are considered dominant for a cluster if the majority of individuals in the cluster had that characteristic. For ordinal variables where a single level does not dominate (all levels contain <50% of individuals), the dominant characteristic is the level for which the majority of individuals in the cluster reported that level or worse. Only characteristics that differed for at least one cluster are shown. A blank cell indicates that the characteristic does not apply to the cluster.

Table 60: Dominant characteristics for gross upper-body movement clusters.

Cluster #	User Characteristic	PAM			HC			GMM		
		2	3	4	2	3	4	2	3	4
1	Age	OA	OA	OA	OA	OA	OA	OA	OA	OA
	Physical Dependence				PD					
	Physical Activity	SD	SD	SD	SD	SD	SD	SD	SD	SD
	Has Arthritis	X			X					
	Has Neck/Back Problems									
	Has Obesity				X	X	X	X		X
	# Cardio-pulmonary Conditions									
2	Age	OA	OA	OA	OA	OA	OA	OA	OA	OA
	Physical Dependence	PD			PD	PD	PD	PD	PD	PD
	Physical Activity	SD	SA	SD	SD	SD	SD	SD	SD	SD
	Has Arthritis	X	X		X	X	X	X	X	X
	Has Neck/Back Problems	X			X	X	X	X	X	
	Has Obesity	X	X	X	X	X	X	X	X	
	# Cardio-pulmonary Conditions	1+			1+	1+	1+	1+		
3	Age		OA	OA		OA	OA		OA	A
	Physical Dependence		PD	PD		PD	PD		PD	PD
	Physical Activity		SD	SD		SD	SD		SD	SA
	Has Arthritis		X	X		X	X		X	X
	Has Neck/Back Problems		X			X	X		X	X
	Has Obesity		X	X		X	X		X	X
	# Cardio-pulmonary Conditions		2+	1+		2+	2+		3+	
4	Age			OA			OA			OA
	Physical Dependence			PD			PD			PD
	Physical Activity			SD			SD			SD
	Has Arthritis			X			X			X
	Has Neck/Back Problems			X						X
	Has Obesity			X			X			X
	# Cardio-pulmonary Conditions			1+			3+			3+

A = Adult, OA = Older Adult, PD = Partial Dependence, SA = Semi-active, SD = Sedentary, X indicates presence of condition.

Table 61: Dominant characteristics for fine motor movement clusters.

Cluster #	User Characteristic	PAM			HC			GMM		
		2	3	4	2	3*	4*	2	3	4
1	Age	OA	OA	OA	OA	OA	OA	OA	A	A
	Physical Dependence									
	Physical Activity	SD	SD	SD	SD	SD	SD	SD	SD	SD
	Has Arthritis							X	X	X
	Has Bone/Joint Injury									
	# Cardio-pulmonary Conditions									
2	Age	OA	OA	OA	OA	OA	OA	OA	OA	OA
	Physical Dependence							PD		
	Physical Activity	SD	A	A	SD	SD	SD	SD	SD	SD
	Has Arthritis	X	X	X	X	X	X	X	X	X
	Has Bone/Joint Injury									
	# Cardio-pulmonary Conditions									
3	Age		OA	OA		A	OA		OA	OA
	Physical Dependence					CD	PD		PD	
	Physical Activity		SD	SD		SD	SD		SD	SA
	Has Arthritis		X	X		X	X		X	X
	Has Bone/Joint Injury									
	# Cardio-pulmonary Conditions					1+			1+	1+
4	Age			OA			OA			OA
	Physical Dependence						CD			PD
	Physical Activity			SD			SD			SD
	Has Arthritis			X			X			X
	Has Bone/Joint Injury			X						
	# Cardio-pulmonary Conditions						1+			

A = Adult, OA = Older Adult, PD = Partial Dependence, CD = Complete Dependence, SA = Semi-active, SD = Sedentary, X indicates presence of condition. *Clustering includes a 1-individual cluster and was eliminated from consideration.

Table 62: Dominant characteristics for cognitive task clusters.

Cluster #	User Characteristic	**PAM**			**HC**			**GMM**		
		2	3	4	2	3	4	2	3	4
1	Age	OA	OA	OA	OA	OA	OA	OA	OA	OA
	Low Energy*				SVD	SVD				
	Cognitive Function									
	Anxiety Frequency	R			R	R	R			
	Anxiety Severity	MLD			MDR	MDR	MLD			
	Depression Frequency				R	R				
	Depression Severity				MDR	MDR				
2	Age	OA	OA	OA	OA	OA	OA	OA	OA	OA
	Low Energy*	SVD	SVD	SVD	SVD	SVD	SVD	SVD	SVD	SVD
	Cognitive Function				I					
	Anxiety Frequency	W	M	R	R	W	W	M	R	R
	Anxiety Severity	MDR	MDR	MLD	MDR	MDR	MDR	MDR	MDR	MLD
	Depression Frequency	W	R	R	R	M	M	R	R	R
	Depression Severity	MDR	MDR	MLD	MLD	MDR	MDR	MDR	MLD	MLD
3	Age		A	A		OA	OA		A	OA
	Low Energy*		MTH	SVD					MTH	SVD
	Cognitive Function		I			I	I		I	I
	Anxiety Frequency		D	W		R	R		D	M
	Anxiety Severity		MDR	MDR		MLD	MLD		MDR	MDR
	Depression Frequency		W	M					D	R
	Depression Severity		MDR	MDR					MDR	MDR
4	Age			A			A			A
	Low Energy*			MTH			MTH			MTH
	Cognitive Function			I			I			I
	Anxiety Frequency			D			D			D
	Anxiety Severity			MDR			MDR			MDR
	Depression Frequency			W			W			D
	Depression Severity			MDR			MDR			SVR

Table 62 (continued)
A = Adult, OA = Older Adult, SVD = Several days, MTH = More than half, R = Rarely, M = Monthly, W = Weekly, D = Daily, MLD = Mild, MDR = Moderate, SVR = Severe, I = Impaired. *Reported as frequency over a period of two weeks.

Appendix D: Student Exercise Case Study Materials

D.1 Case Study 1 Material: Mobility Device

Problem Statement: The most common disabilities in the United States are mobility related. Many devices exist to aid mobility, however due to the highly varied characteristics and needs of mobility impaired individuals, not one size fits all.

For this task, you will identify functions for a product to aid mobility in individuals with serious difficulty walking. This device should be for usable in, but not limited to, the home and outside on paved surfaces. You will complete two different exercises concerning this design problem. You should try to balance product accessibility with product complexity. Summary statistics for the mobility-impaired population have been provided to help you perform this task.

Mobility-impaired* Population Summary Statistics:

*Defined as individuals who report serious difficulty walking or climbing stairs

- **Median Age:** 65.0
- **Gender:** 55.9% Male, 44.1% Female
- **Education:**
    - Less than high school: 30.0%
    - High school: 26.8%
    - Some college or associate degree: 29.7%
    - College degree: 12.8%
- **Uses equipment to walk:** 63.1%
- **Difficulty walking for a quarter mile**

- At least *some* difficulty: 73.2%
- Significant difficulty or cannot do: 30.9%

- **Reports difficulty dressing and bathing, reaching up, or moving large objects:**
  - At least *some* difficulty: 89.0%
  - Significant difficulty or cannot do: 58.2%
- **Reports difficulty using silverware or grasping/moving small objects:**
  - At least *some* difficulty: 47.9%
  - Significant difficulty or cannot do: 14.2%
- **Recreational or work activity (weekly):**
  - Neither: 54.0%
  - Moderate: 25.2%
  - Vigorous: 20.8%
- **Has arthritis, gout, bone/joint injury, or back/neck problem:**
  - 1 or more conditions: 84.8%
  - 2 or more conditions: 54.0%
  - 3 or more conditions: 16.3%
- **Cardiovascular conditions (e.g. congestive heart failure, coronary heart disease):**
  - 1 or more conditions: 23.0%
  - 2 or more conditions: 7.9%
  - 3 or more conditions: 1.3%
- **Pulmonary conditions (e.g. asthma, emphysema):**
  - 1 or more conditions: 37.6%
  - 2 or more conditions: 16.4%

    - 3 or more conditions: 7.4%
- **Obesity:** 51.8%
- **Stroke:** 13.9%
- **Difficulty hearing:** 23.9%
- **Difficulty seeing:** 20.5%
- **Reports difficulty managing money:**
    - At least *some* difficulty: 33.1%
    - Significant difficulty or cannot do: 15.7%
- **Reports confusion/memory problems or difficulty concentrating and making decisions:** 40.4%
- **Reports low energy levels:**
    - Several days a week or more: 69.6%
    - Nearly every day: 21.8%
- **Anxiety (frequency):**
    - Monthly or more: 53.2%
    - Weekly or more: 41.4%
    - Daily: 27.1%
- **Anxiety (severity):**
    - Mild: 36.5%
    - Moderate: 43.1%
    - Severe: 20.2%
- **Depression (frequency)**
    - Monthly or more: 38.7%

    - Weekly or more: 27.6%
    - Daily: 14.6%
- **Depression (severity)**
    - Mild: 36.7%
    - Moderate: 38.5%
    - Severe: 24.4%

D.2 Case Study 2 Material: Glucose Monitoring Device

<u>Problem Statement:</u> For individuals with type 1 and type 2 diabetes, monitoring blood glucose levels is critical for successfully managing their disease. Many diagnostic devices exist for the purpose, typically in the form of a handheld device. Despite the availability of these devices, successful self-monitoring of blood glucose levels remains low.

For this task, a set of system functions to satisfy this problem has been provided. You are asked to develop a family of blood glucose monitoring devices that caters to the capabilities of the diabetes population. To do so, you will assign the given functions to human or machine for several product family members. You should try to balance product accessibility with product complexity. Summary statistics for the diabetes population have been provided to help you perform this task.

In most cases, these systems work by extracting some glucose carrying medium from the body. This medium is most often blood but can also be urine or saliva. The medium undergoes some change (e.g. a chemical reaction), and from that reaction the concentration of glucose can be derived. This value can then be used to determine if the user is in a healthy state.

Now knowing the basic processes that must happen to satisfy this task, the following are baseline functions for the case study:

1. **Sense** the state of system functions
2. **Process** the sensory information
3. **Store** the sensory information
4. **Extract** blood glucose from the body
5. **Transfer** blood glucose
6. **Store** blood glucose
7. **Condition** blood glucose
8. **Indicate** glucose levels
9. **Detect** glucose levels
10. **Process** glucose levels

<u>Diabetes Population Summary Statistics:</u>

- **Median Age:** 65.0
- **Gender:** 54.2% Male, 45.8% Female
- **Education:**
    - Less than high school: 27.0%
    - High school: 22.6%
    - Some college or associate degree: 29.9%
    - College degree: 18.4%

- **Reports difficulty dressing and bathing, reaching up, or moving large objects:**
    - At least *some* difficulty: 62.2%
    - Significant difficulty or cannot do: 33.6%
- **Reports difficulty using silverware or grasping/moving small objects:**
    - At least *some* difficulty: 30.3%
    - Significant difficulty or cannot do: 6.9%
- **Recreational or work activity (weekly):**
    - Neither: 45.7%
    - Moderate: 30.8%
    - Vigorous: 23.5%
- **Has arthritis, gout, bone/joint injury, or back/neck problem:**
    - 1 or more conditions: 65.4%
    - 2 or more conditions: 34.7%
    - 3 or more conditions: 9.1%
- **Cardiovascular conditions (e.g. congestive heart failure, coronary heart disease):**
    - 1 or more conditions: 22.0%
    - 2 or more conditions: 8.7%
    - 3 or more conditions: 2.3%
- **Pulmonary conditions (e.g. asthma, emphysema):**
    - 1 or more conditions: 29.5%
    - 2 or more conditions: 10.3%

    - 3 or more conditions: 3.8%
- **Obesity:** 59.0%
- **Had a stroke:** 10.9%
- **Difficulty hearing:** 19.5%
- **Difficulty seeing:** 14.9%
- **Reports difficulty managing money:**
    - At least *some* difficulty: 21.4%
    - Significant difficulty or cannot do: 10.0%
- **Reports confusion/memory problems or difficulty concentrating and making decisions:** 23.7%
- **Reports low energy levels:**
    - Several days a week or more: 52.5%
    - Nearly every day: 11.8%
- **Anxiety (frequency):**
    - Monthly or more: 38.1%
    - Weekly or more: 26.0%
    - Daily: 15.2%
- **Anxiety (severity):**
    - Mild: 45.9%
    - Moderate: 39.3%
    - Severe: 14.8%
- **Depression (frequency)**
    - Monthly or more: 23.4%

- Weekly or more: 16.1%
- Daily: 7.8%

- **Depression (severity)**
  - Mild: 42.5%
  - Moderate: 38.2%
  - Severe: 19.2%

Bibliography

Addae, J. H., Sun, X., Towey, D., & Radenkovic, M. (2019). Exploring user behavioral data for adaptive cybersecurity. *User Modeling and User-Adapted Interaction*, *29*(3), 701–750. https://doi.org/10.1007/s11257-019-09236-5

Agnisarman, S. O., Chalil Madathil, K., Smith, K., Ashok, A., Welch, B., & McElligott, J. T. (2017). Lessons learned from the usability assessment of home-based telemedicine systems. *Applied Ergonomics*, *58*, 424–434. https://doi.org/10.1016/j.apergo.2016.08.003

Ahlstrom, V. (2016). *Human Factors Design Standard (DOT/FAA/HF-STD-001B)*.

Ahmad, A., & Khan, S. S. (2019). Survey of State-of-the-Art Mixed Data Clustering Algorithms. *IEEE Access*, *7*, 31883–31902. https://doi.org/10.1109/ACCESS.2019.2903568

Akai, R., Amaya, H., & Fujita, K. (2010). *Product Family Deployment Through Optimal Resource Allocation Under Market System*. 277–290. https://doi.org/10.1115/DETC2010-28662

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low Choy, S., Mengersen, K., & Rousseau, J. (2012). Combining expert opinions in prior elicitation. *Bayesian Analysis*, *7*(3), 503–532.

Alessa, T., Hawley, M. S., Hock, E. S., & Witte, L. de. (2019). Smartphone Apps to Support Self-Management of Hypertension: Review and Content Analysis. *JMIR MHealth and UHealth*, *7*(5), e13645. https://doi.org/10.2196/13645

Alizadeh, M., Esfahani, M. N., Tian, W., & Ma, J. (2019). Data-Driven Energy Efficiency and Part Geometric Accuracy Modeling and Optimization of Green

Fused Filament Fabrication Processes. *Journal of Mechanical Design*, *142*(041701). https://doi.org/10.1115/1.4044596

Allsworth, J. E. (2015). Invited Commentary: Recruiting for Epidemiologic Studies Using Social Media. *American Journal of Epidemiology*, *181*(10), 747–749. https://doi.org/10.1093/aje/kwv007

Alsager Alzayed, M., McComb, C., Menold, J., Huff, J., & Miller, S. R. (2020). Are you feeling me? An exploration of empathy development in engineering design education. *Journal of Mechanical Design*, 1–57. https://doi.org/10.1115/1.4048624

An, J., Kwak, H., & Jansen, B. J. (2017). Personas for Content Creators via Decomposed Aggregate Audience Statistics. *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 632–635.

Andriella, A., Alenyà, G., Hernández-Farigola, J., & Torras, C. (2018). Deciding the different robot roles for patient cognitive training. *International Journal of Human-Computer Studies*, *117*, 20–29. https://doi.org/10.1016/j.ijhcs.2018.03.004

Anvari, F., Richards, D., Hitchens, M., Babar, M. A., Tran, H. M. T., & Busch, P. (2017). An empirical investigation of the influence of persona with personality traits on conceptual design. *Journal of Systems and Software*, *134*, 324–339. https://doi.org/10.1016/j.jss.2017.09.020

Asche, C., LaFleur, J., & Conner, C. (2011). A Review of Diabetes Treatment Adherence and the Association with Clinical and Economic Outcomes.

Clinical Therapeutics, *33*, 74–109. https://doi.org/10.1016/j.clinthera.2011.01.019

August, T., Dao, D., & Kim, K. (2019). Market Segmentation and Software Security: Pricing Patching Rights. *Management Science*. https://doi.org/10.1287/mnsc.2018.3153

Aycock, K. I., De Farias, J. G., Kim, J.-E., Shi, C., & Parkinson, M. (2015). *A design for human variability (DfHV) approach to predicting appropriate dimensions for pediatric-sized prosthetic heart valves*.

Baker, N. A., & Rogers, J. C. (2010). Association between computer use speed and age, impairments in function, and touch typing training in people with rheumatoid arthritis. *Arthritis Care & Research*, *62*(2), 242–250. https://doi.org/10.1002/acr.20074

Barnes, M. J. (2014). Alcohol: Impact on Sports Performance and Recovery in Male Athletes. *Sports Medicine*, *44*(7), 909–919. https://doi.org/10.1007/s40279-014-0192-8

Barnum, C. M. (2020). *Usability Testing Essentials: Ready, Set ...Test!* (2nd ed.). Morgan Kaufmann.

Bekhradi, A., Yannou, B., Farel, R., Zimmer, B., & Chandra, J. (2015). Usefulness Simulation of Design Concepts. *Journal of Mechanical Design*, *137*(7), 071414-071414–12. https://doi.org/10.1115/1.4030180

Bindewald, J. M., Miller, M. E., & Peterson, G. L. (2014). A function-to-task process model for adaptive automation system design. *International Journal of*

Human-Computer Studies, *72*(12), 822–834. https://doi.org/10.1016/j.ijhcs.2014.07.004

Blackwell, D. L., Lucas, J. W., & Clarke, T. C. (2014). Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2012. *Vital and Health Statistics. Series 10, Data from the National Health Survey*, *10*(260), 1–161.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. D. McKay.

Bose, A., Munir, A., & Shabani, N. (2020). A Quantitative Analysis of Big Data Clustering Algorithms for Market Segmentation in Hospitality Industry. *2020 IEEE International Conference on Consumer Electronics (ICCE)*, 1–6. https://doi.org/10.1109/ICCE46568.2020.9043023

Boyd, D. K., & Parkinson, M. B. (2015). Using Multivariate Analysis to Select Accommodation Boundary Manikins From a Population Database. *DAC 2015*, V02BT03A021. https://doi.org/10.1115/DETC2015-47504

Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statistical Software*, *25*(1), 1–22. https://doi.org/10.18637/jss.v025.i04

Brolin, E., Högberg, D., Hanson, L., & Örtengren, R. (2016). Generation and evaluation of distributed cases by clustering of diverse anthropometric data. *International Journal of Human Factors Modelling and Simulation*, *5*(3), 210–229. https://doi.org/10.1504/IJHFMS.2016.079706

Brolin, E., Högberg, D., Hanson, L., & Örtengren, R. (2017). Adaptive regression model for synthesizing anthropometric population data. *International Journal* *of Industrial Ergonomics*, *59*(Supplement C), 46–53. https://doi.org/10.1016/j.ergon.2017.03.008

Brownstein, N. C., Louis, T. A., O'Hagan, A., & Pendergast, J. (2019). The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision-Making. *The American Statistician*, *73*(sup1), 56–68. https://doi.org/10.1080/00031305.2018.1529623

Cage, K. (2017). *Optimizing Mass Customization Through Interaction Variability and Manufacturing Trade-offs* [University of Maryland]. http://drum.lib.umd.edu/handle/1903/20813

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Camus, K. A. (2015). *Once Careless, Always Careless? Temporal and Situational Stability of Insufficient Effort Responding (IER)* [Wright State University]. https://corescholar.libraries.wright.edu/etd_all/1617

Carey, M., White, E. J., McMahon, M., & O'Sullivan, L. W. (2019). Using personas to exploit environmental attitudes and behaviour in sustainable product design. *Applied Ergonomics*, *78*, 97–109. https://doi.org/10.1016/j.apergo.2019.02.005

Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, *10*(6), 807–815. https://doi.org/10.1016/j.cptl.2018.03.019

Centers for Disease Control and Prevention. (2018). *National Health Interview Survey*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nhis/index.htm

Centers for Disease Control and Prevention. (2019). *National Health and Nutrition Examination Survey Data*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nhanes/index.htm

Centers for Disease Control and Prevention. (2020a). *National Diabetes Statistics Report*. Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services. https://www.cdc.gov/diabetes/data/statistics-report/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fdiabetes%2Fdata%2Fstatistics%2Fstatistics-report.html

Centers for Disease Control and Prevention. (2020b, September 17). *Physical Activity Recommendations By Age Group*. Centers for Disease Control and Prevention. https://www.cdc.gov/physicalactivity/basics/age-chart.html

Chaniaud, N., Métayer, N., Megalakaki, O., & Loup-Escande, E. (2020). Effect of Prior Health Knowledge on the Usability of Two Home Medical Devices: Usability Study. *JMIR MHealth and UHealth*, *8*(9), e17983. https://doi.org/10.2196/17983

Chatterjee, T., Chakraborty, S., & Chowdhury, R. (2019). A Critical Review of Surrogate Assisted Robust Design Optimization. *Archives of Computational Methods in Engineering*, *26*(1), 245–274. https://doi.org/10.1007/s11831-017-9240-5

Chomutare, T., Fernandez-Luque, L., Årsand, E., & Hartvigsen, G. (2011). *Features of Mobile Diabetes Applications: Review of the Literature and Analysis of Current Applications Compared Against Evidence-Based Guidelines*. https://doi.org/10.2196/jmir.1874

Christensen, T., Riis, A. H., Hatch, E. E., Wise, L. A., Nielsen, M. G., Rothman, K. J., Sørensen, H. T., & Mikkelsen, E. M. (2017). Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. *Journal of Medical Internet Research*, *19*(3), e58. https://doi.org/10.2196/jmir.6716

Ciavola, B. T., Wu, C., & Gershenson, J. K. (2015). Integrating Function- and Affordance-Based Design Representations. *Journal of Mechanical Design*, *137*(5), 051101. https://doi.org/10.1115/1.4029519

Cieza, A., Stucki, A., Geyh, S., Berteanu, M., Quittan, M., Simon, A., Kostanjsek, N., Stucki, G., & Walsh, N. (2004). ICF Core Sets for chronic ischaemic heart disease. *Journal of Rehabilitation Medicine*, *36*, 94–99.

Cifter, A. S. (2017). Blood Pressure Monitor Usability Problems Detected Through Human Factors Evaluation. *Ergonomics in Design*, *25*(3), 11–19. https://doi.org/10.1177/1064804617691397

Çığ Karaman, Ç., & Sezgin, T. M. (2018). Gaze-based predictive user interfaces: Visualizing user intentions in the presence of uncertainty. *International Journal of Human-Computer Studies*, *111*, 78–91. https://doi.org/10.1016/j.ijhcs.2017.11.005

Claypoole, V. L., Dever, D. A., Denues, K. L., & Szalma, J. L. (2019). The Effects of Event Rate on a Cognitive Vigilance Task. *Human Factors*, *61*(3), 440–450. https://doi.org/10.1177/0018720818790840

Colson, A. R., & Cooke, R. M. (2017). Cross validation for the classical model of structured expert judgment. *Reliability Engineering & System Safety*, *163*, 109–120. https://doi.org/10.1016/j.ress.2017.02.003

Colson, A. R., & Cooke, R. M. (2018). Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. *Review of Environmental Economics and Policy*, *12*(1), 113–132. https://doi.org/10.1093/reep/rex022

Cooke, R. M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press.

Cormier, P., Olewnik, A., & Lewis, K. (2014). Toward a formalization of affordance modeling for engineering design. *Research in Engineering Design*, *25*(3), 259–277. https://doi.org/10.1007/s00163-014-0179-3

Crompton, H., Burke, D., & Lin, Y.-C. (2019). Mobile learning and student cognition: A systematic review of PK-12 research using Bloom's Taxonomy. *British Journal of Educational Technology*, *50*(2), 684–701. https://doi.org/10.1111/bjet.12674

Cuneo, A., Traverso, A., & Shahpar, S. (2017, August 17). *Comparative Analysis of Methodologies for Uncertainty Propagation and Quantification*. ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition. https://doi.org/10.1115/GT2017-63238

Czaja, S. J., Zarcadoolas, C., Vaughon, W. L., Lee, C. C., Rockoff, M. L., & Levy, J. (2015). The Usability of Electronic Personal Health Record Systems for an Underserved Adult Population. *Human Factors*, *57*(3), 491–506. https://doi.org/10.1177/0018720814549238

Dascalu, M., Moldoveanu, A., Balan, O., Lupu, R. G., Ungureanu, F., & Caraiman, S. (2017). Usability assessment of assistive technology for blind and visually impaired. *2017 E-Health and Bioengineering Conference (EHB)*, 523–526. https://doi.org/10.1109/EHB.2017.7995476

DAU. (2020). *Defense Acquisition Guidebook (DAG)*. Defense Acquisition University (DAU)/U.S. Department of Defense (DoD). https://www.dau.edu/tools/dag

de Vries, C., & Parkinson, M. B. (2014). Limiting disproportionate disaccommodation in design for human variability. *Ergonomics*, *57*(1), 52–65. https://doi.org/10.1080/00140139.2013.853102

de Winter, J. C. F., & Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work*, *16*(1), 1–11. https://doi.org/10.1007/s10111-011-0188-1

de Winter, J. C. F., & Hancock, P. A. (2015). Reflections on the 1951 Fitts List: Do Humans Believe Now that Machines Surpass them? *Procedia Manufacturing*, *3*, 5334–5341. https://doi.org/10.1016/j.promfg.2015.07.641

Deb, K. (2009). *Multi-Objective Optimization Using Evolutionary Algorithms* (1st edition). Wiley.

Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or Abracadabra? Progress on Human–Automation Co-ordination. *Cognition, Technology & Work*, *4*(4), 240–244. https://doi.org/10.1007/s101110200022

Del Missier, F., Mäntylä, T., & Nilsson, L.-G. (2015). Aging, Memory, and Decision Making. In T. M. Hess, J. Strough, & C. E. Löckenhoff (Eds.), *Aging and Decision Making* (pp. 127–148). Academic Press. https://doi.org/10.1016/B978-0-12-417148-0.00007-8

Delliaux, S., Delaforge, A., Deharo, J.-C., & Chaumet, G. (2019). Mental Workload Alters Heart Rate Variability, Lowering Non-linear Dynamics. *Frontiers in Physiology*, *10*. https://doi.org/10.3389/fphys.2019.00565

Department of Defense. (2016). *Human Engineering Requirements for Military Systems, Equipment, and Facilities (MIL-STD-46855A)*.

Dias, L. C., Morton, A., & Quigley, J. (Eds.). (2018). *Elicitation: The Science and Art of Structuring Judgement*. Springer International Publishing. https://www.springer.com/us/book/9783319650517

Dindar, S., Kaewunruen, S., & An, M. (2020). Bayesian network-based human error reliability assessment of derailments. *Reliability Engineering & System Safety*, *197*, 106825. https://doi.org/10.1016/j.ress.2020.106825

Domańska, B., VanLunen, B., Peterson, L., Mountian, I., & Schiff, M. (2017). Comparative usability study for a certolizumab pegol autoinjection device in patients with rheumatoid arthritis. *Expert Opinion on Drug Delivery*, *14*(1), 15–22. https://doi.org/10.1080/17425247.2016.1256283

Donndelinger, J., & Ferguson, S. M. (2017). *Design for Marketing Mix: The Past, Present, and Future of Market-Driven Product Design*. V02AT03A040. https://doi.org/10.1115/DETC2017-68275

Emerson, P. (2013). The original Borda count and partial voting. *Social Choice and Welfare*, *40*(2), 353–358. https://doi.org/10.1007/s00355-011-0603-9

Endsley, M., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462–492. https://doi.org/10.1080/001401399185595

Fauth, E. B., Schaefer, S. Y., Zarit, S. H., Ernsth-Bravell, M., & Johansson, B. (2017). Associations Between Fine Motor Performance in Activities of Daily Living and Cognitive Ability in a Nondemented Sample of Older Adults: Implications for Geriatric Physical Rehabilitation. *Journal of Aging and Health*, *29*(7), 1144–1159. https://doi.org/10.1177/0898264316654674

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a Characterization of Adaptive Systems: A Framework for Researchers and System Designers. *Human Factors*, *54*(6), 1008–1024. https://doi.org/10.1177/0018720812443983

Feigh, K. M., & Pritchett, A. R. (2014). Requirements for Effective Function Allocation: A Critical Review. *Journal of Cognitive Engineering and Decision Making*, *8*(1), 23–32. https://doi.org/10.1177/1555343413490945

Ferguson, T., Greene, M., Repetti, F., Lewis, K., & Behdad, S. (2015). *Combining Anthropometric Data and Consumer Review Content to Inform Design for Human Variability*. ASME 2015 International Design Engineering Technical

Conferences and Computers and Information in Engineering Conference. https://doi.org/10.1115/DETC2015-47640

Fiorineschi, L., Rissone, P., & Rotini, F. (2014). Modularization vs. Innovation. *International Journal of Innovation Science*, *6*(1), 29–42. https://doi.org/10.1260/1757-2223.6.1.29

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2012). Towards a dynamic balance between humans and automation: Authority, ability, responsibility and control in shared and cooperative control situations. *Cognition, Technology & Work*, *14*(1), 3–18. https://doi.org/10.1007/s10111-011-0191-6

Food and Drug Administration. (2016). *Applying Human Factors and Usability Engineering to Medical Devices Guidance for Industry and Food and Drug Administration Staff* (p. 49) [Guidance]. https://www.fda.gov/media/80481/download

Fujita, K., Amaya, H., & Akai, R. (2013). Mathematical model for simultaneous design of module commonalization and supply chain configuration toward global product family. *Journal of Intelligent Manufacturing*, *24*(5), 991–1004. https://doi.org/10.1007/s10845-012-0641-x

Fuld, R. B. (2000). The fiction of function allocation, revisited. *International Journal of Human-Computer Studies*, *52*(2), 217–233. https://doi.org/10.1006/ijhc.1999.0286

Fung, C. H., Igodan, U., Alessi, C., Martin, J. L., Dzierzewski, J. M., Josephson, K., & Kramer, B. J. (2015). Human factors/usability barriers to home medical devices among individuals with disabling conditions: In-depth interviews with positive airway pressure device users. *Disability and Health Journal*, *8*(1), 86–92. https://doi.org/10.1016/j.dhjo.2014.06.002

Furman, S. M., Theofanos, M. F., & Wald, H. (2014). *Human Engineering Design Criteria Standards Part 3: Interim Steps*. https://www.nist.gov/publications/human-engineering-design-criteria-standards-part-3-interim-steps

Gajawada, S., & Toshniwal, D. (2012). Hybrid Cluster Validation Techniques. In D. C. Wyld, J. Zizka, & D. Nagamalai (Eds.), *Advances in Computer Science, Engineering & Applications* (pp. 267–273). Springer. https://doi.org/10.1007/978-3-642-30111-7_25

Galvao, A. B., & Sato, K. (2005). Affordances in Product Architecture: Linking Technical Functions and Users' Tasks. *Proceedings of IDETC/CIE*, 143–153. https://doi.org/10.1115/DETC2005-84525

Garneau, C. J., Nadadur, G., & Parkinson, M. B. (2014). Considering Human Variability When Implementing Product Platforms. In T. W. Simpson, J. (Roger) Jiao, Z. Siddique, & K. Hölttä-Otto (Eds.), *Advances in Product Family and Product Platform Design* (pp. 559–585). Springer New York. https://doi.org/10.1007/978-1-4614-7937-6_22

Garneau, C. J., & Parkinson, M. B. (2011). A comparison of methodologies for designing for human variability. *Journal of Engineering Design*, *22*(7), 505–521. https://doi.org/10.1080/09544820903535404

Gauss, L., Lacerda, D. P., & Cauchick Miguel, P. A. (2021). Module-based product family design: Systematic literature review and meta-synthesis. *Journal of Intelligent Manufacturing*, *32*(1), 265–312. https://doi.org/10.1007/s10845-020-01572-3

Georgiou, T., & Demiris, Y. (2017). Adaptive user modelling in car racing games using behavioural and physiological data. *User Modeling and User-Adapted Interaction*, *27*(2), 267–311. https://doi.org/10.1007/s11257-017-9192-3

Geyh, S., Cieza, A., Schouten, J., Dickson, H., Frommelt, P., Omar, Z., Kostanjsek, N., Ring, H., & Stucki, G. (2004). ICF Core Sets for stroke. *Journal of Rehabilitation Medicine*, *36*(1), 135–141. https://doi.org/10.1080/16501960410016776

Ghosh, D., Olewnik, A., & Lewis, K. (2016, December 5). *Product "In-Use" Context Identification Using Feature Learning Methods*. ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. https://doi.org/10.1115/DETC2016-59645

Gill, A. S., Summers, J. D., & Turner, C. J. (2017). Comparing function structures and pruned function structures for market price prediction: An approach to benchmarking representation inferencing value. *AI EDAM*, *31*(4), 550–566. https://doi.org/10.1017/S0890060417000543

Godwin, A., Eger, T., Salmoni, A., Grenier, S., & Dunn, P. (2007). Postural implications of obtaining line–of–sight for seated operators of underground mining load-haul-dump vehicles. *Ergonomics*, *50*(2), 192–207. https://doi.org/10.1080/00140130600951970

Goel, A. K., Rugaber, S., & Vattam, S. (2009). Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language. *AI EDAM*, *23*(1), 23–35. https://doi.org/10.1017/S0890060409000080

Gordon, C. C., Churchill, T., Clauser, C. E., Bradtmiller, B., & McConville, J. T. (1989). *Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics 1988*. Anthropology Research Project Inc. http://www.dtic.mil/docs/citations/ADA225094

Gosling, J. P. (2018). SHELF: The Sheffield Elicitation Framework. In L. C. Dias, A. Morton, & J. Quigley (Eds.), *Elicitation: The Science and Art of Structuring Judgement* (pp. 61–93). Springer International Publishing. https://doi.org/10.1007/978-3-319-65052-4_4

Greenwood, D. A., Gee, P. M., Fatkin, K. J., & Peeples, M. (2017). A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. *Journal of Diabetes Science and Technology*, *11*(5), 1015–1027. https://doi.org/10.1177/1932296817713506

Guilera, G., Pino, O., Barrios, M., Rojo, E., Vieta, E., & Gómez-Benito, J. (2020). Towards an ICF Core Set for functioning assessment in severe mental disorders: Commonalities in bipolar disorder, depression and schizophrenia. *Psicothema*, *32*(1), 7–14. https://doi.org/10.7334/psicothema2019.186

Hamid, R., Baharom, S., Hamzah, N., Badaruzzaman, W. H. W., Rahmat, R. A. O. K., & RaihanTaha, M. (2012). Assessment of Psychomotor Domain in Materials Technology Laboratory Work. *Procedia - Social and Behavioral Sciences*, *56*, 718–723. https://doi.org/10.1016/j.sbspro.2012.09.708

Hancock, P. A., & Scallen, S. F. (1996). The Future of Function Allocation. *Ergonomics in Design*, *4*(4), 24–29. https://doi.org/10.1177/106480469600400406

Handl, J., Knowles, J., & Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, *21*(15), 3201–3212. https://doi.org/10.1093/bioinformatics/bti517

Hanea, A. M., & Nane, G. F. (2019). Calibrating experts' probabilistic assessments for improved probabilistic predictions. *Safety Science*, *118*, 763–771. https://doi.org/10.1016/j.ssci.2019.05.048

Harding, J. L., Pavkov, M. E., Magliano, D. J., Shaw, J. E., & Gregg, E. W. (2019). Global trends in diabetes complications: A review of current evidence. *Diabetologia*, *62*(1), 3–16. https://doi.org/10.1007/s00125-018-4711-2

Harrow, A. J. (1972). *A taxonomy of the psychomotor domain: A guide for developing behavioral objectives*. D. McKay Co. https://books.google.com/books?id=at87AAAAIAAJ

Harvey, A. J., Bayless, S. J., & Hyams, G. (2018). Alcohol increases inattentional blindness when cognitive resources are not consumed by ongoing task demands. *Psychopharmacology*, *235*(1), 309–315. https://doi.org/10.1007/s00213-017-4772-9

He, Z., Yen, G. G., & Yi, Z. (2019). Robust Multiobjective Optimization via Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, *23*(2), 316–330. https://doi.org/10.1109/TEVC.2018.2859638

Heinemann, L., Drossel, D., Freckmann, G., & Kulzer, B. (2016). Usability of Medical Devices for Patients With Diabetes Who Are Visually Impaired or Blind. *Journal of Diabetes Science and Technology*, *10*(6), 1382–1387. https://doi.org/10.1177/1932296816666536

Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution, 9*(1), 169–180. https://doi.org/10.1111/2041-210X.12857

Heymsfield, S. B., Bourgeois, B., Ng, B. K., Sommer, M. J., Li, X., & Shepherd, J. A. (2018). Digital anthropometry: A critical review. *European Journal of Clinical Nutrition*, *72*(5), 680–687. https://doi.org/10.1038/s41430-018-0145-7

Hirtz, J., Stone, R. B., McAdams, D. A., Szykman, S., & Wood, K. L. (2002). A functional basis for engineering design: Reconciling and evolving previous efforts. *Research in Engineering Design*, *13*(2), 65–82. https://doi.org/10.1007/s00163-001-0008-3

Idoughi, D., Seffah, A., & Kolski, C. (2012). Adding user experience into the interactive service design loop: A persona-based approach. *Behaviour & Information Technology*, *31*(3), 287–303. https://doi.org/10.1080/0144929X.2011.563799

Ilbeygi, M., Kangavari, M. R., & Golmohammadi, S. A. (2019). Equipping the ACT-R cognitive architecture with a temporal ratio model of memory and using it in a new intelligent adaptive interface. *User Modeling and User-Adapted Interaction*, *29*(5), 943–976. https://doi.org/10.1007/s11257-019-09239-2

Irshad, L., Ahmed, S., Demirel, O., & Tumer, I. Y. (2018, November 2). *Identification of Human Errors During Early Design Stage Functional Failure Analysis*. ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. https://doi.org/10.1115/DETC2018-85979

Irshad, L., Ahmed, S., Demirel, O., & Tumer, I. Y. (2019, January 6). Coupling Digital Human Modeling with Early Design Stage Human Error Analysis to Assess Ergonomic Vulnerabilities. *AIAA Scitech 2019 Forum*. https://doi.org/10.2514/6.2019-2349

Jeon, Y., Jeon, S. G., & Han, K. (2020). Better targeting of consumers: Modeling multifactorial gender and biological sex from Instagram posts. *User Modeling and User-Adapted Interaction*. https://doi.org/10.1007/s11257-020-09260-w

Jevtić, A., Valle, A. F., Alenyà, G., Chance, G., Caleb-Solly, P., Dogramadzi, S., & Torras, C. (2019). Personalized Robot Assistant for Support in Dressing. *IEEE Transactions on Cognitive and Developmental Systems*, *11*(3), 363–374. https://doi.org/10.1109/TCDS.2018.2817283

Jiao, J. (Roger), Simpson, T. W., & Siddique, Z. (2007). Product family design and platform-based product development: A state-of-the-art review. *Journal of Intelligent Manufacturing*, *18*(1), 5–29. https://doi.org/10.1007/s10845-007-0003-2

Johansen, E. (2018). Making Human Factors Affordable for Medical Device and Global Health Startups. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, *7*(1), 140–147. https://doi.org/10.1177/2327857918071036

Kaber, D. B. (2018). Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 7–24. https://doi.org/10.1177/1555343417737203

Kalpokaite, N., & Radivojevic, I. (2019). Demystifying Qualitative Data Analysis for Novice Qualitative Researchers. *The Qualitative Report*, *24*(13). /paper/Demystifying-Qualitative-Data-Analysis-for-Novice-Kalpokaite-Radivojevic/9d4b0f14961993b1cd7d71a41d3a9a63b3d179b5

Kannry, J., Beuria, P., Wang, E., & Nissim, J. (2012). Personal health records: Meaningful use, but for whom? *The Mount Sinai Journal of Medicine, New York*, *79*(5), 593–602. https://doi.org/10.1002/msj.21334

Kidwell, B., Calhoun, G. L., Ruff, H. A., Parasuraman, R., & Mason, G. (2012). *Adaptable and Adaptive Automation for Supervisory Control of Multiple Autonomous Vehicles*.

Kim, S., & Moon, S. K. (2017). Sustainable product family configuration based on a platform strategy. *Journal of Engineering Design*, *28*(10–12), 731–764. https://doi.org/10.1080/09544828.2017.1393657

Knisely, B. M., Joyner, J. S., Rutkowski, A. M., Wong, M., Barksdale, S., Hotham, H., Kharod, K., & Vaughn-Cooke, M. (2020). A Cognitive Decomposition to

Empirically Study Human Performance in Control Room Environments. *International Journal of Human-Computer Studies.* https://doi.org/10.1016/j.ijhcs.2020.102438

Knisely, B. M., Levine, C., Kharod, K. C., & Vaughn-Cooke, M. (2020). An Analysis of FDA Adverse Event Reporting Data for Trends in Medical Device Use Error. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, *9*, 130–134. https://doi.org/10.1177/2327857920091024

Knisely, B. M., & Vaughn-Cooke, M. (2020a). Virtual Modeling of User Populations and Formative Design Parameters. *Systems*, *8*(4), 35. https://doi.org/10.3390/systems8040035

Knisely, B. M., & Vaughn-Cooke, M. (2020b). A Comprehensive Survey of Preference for Personal Health Record Features in Patients with Diabetes. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, *9*, 171–175. https://doi.org/10.1177/2327857920091025

Knol, A. B., Slottje, P., van der Sluijs, J. P., & Lebret, E. (2010). The use of expert elicitation in environmental health impact assessment: A seven step procedure. *Environmental Health: A Global Access Science Source*, *9*, 19. https://doi.org/10.1186/1476-069X-9-19

Kong, J., Zhang, W. Y., Yu, N., & Xia, X. J. (2011). Design of human-centric adaptive multimodal interfaces. *International Journal of Human-Computer Studies*, *69*(12), 854–869. https://doi.org/10.1016/j.ijhcs.2011.07.006

Korsgaard, D., Bjørner, T., Sørensen, P. K., & Burelli, P. (2020). Creating user stereotypes for persona development from qualitative data through semi-automatic subspace clustering. *User Modeling and User-Adapted Interaction*, *30*(1), 81–125. https://doi.org/10.1007/s11257-019-09252-5

Kota, S., & Chakrabarti, A. (2010). A method for Estimating the Degree of Uncertainty With Respect to Life Cycle Assessment During Design. *Journal of Mechanical Design*, *132*(9). https://doi.org/10.1115/1.4002163

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Lalwani, A., & Agrawal, S. (2018). Validating Revised Bloom's Taxonomy Using Deep Knowledge Tracing. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), *Artificial Intelligence in Education* (pp. 225–238). Springer International Publishing. https://doi.org/10.1007/978-3-319-93843-1_17

Lavie, T., & Meyer, J. (2010). Benefits and costs of adaptive user interfaces. *International Journal of Human-Computer Studies*, *68*(8), 508–524. https://doi.org/10.1016/j.ijhcs.2010.01.004

Leonard, V. K., Jacko, J. A., Yi, J. S., & Sainfort, F. (2006). Human Factors and Ergonomic Methods. In G. S. P. D. C. Department essor, and Head of the (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 292–321). John Wiley & Sons, Inc. https://doi.org/10.1002/0470048204.ch11

Li, L. C., Adam, P. M., Townsend, A. F., Lacaille, D., Yousefi, C., Stacey, D., Gromala, D., Shaw, C. D., Tugwell, P., & Backman, C. L. (2013). Usability testing of ANSWER: A web-based methotrexate decision aid for patients with rheumatoid arthritis. *BMC Medical Informatics and Decision Making*, *13*(1), 131. https://doi.org/10.1186/1472-6947-13-131

Li, M., Chen, T., & Yao, X. (2020). How to Evaluate Solutions in Pareto-based Search-Based Software Engineering? A Critical Review and Methodological Guidance. *ArXiv:2002.09040 [Cs]*. https://doi.org/10.1109/TSE.2020.3036108

Liao, H., Groth, K., & Stevens-Adams, S. (2015). Challenges in leveraging existing human performance data for quantifying the IDHEAS HRA method. *Reliability Engineering & System Safety*, *144*, 159–169. https://doi.org/10.1016/j.ress.2015.07.018

Lichtendahl, K. C., Grushka-Cockayne, Y., & Winkler, R. L. (2013). Is It Better to Average Probabilities or Quantiles? *Management Science*, *59*(7), 1594–1611. https://doi.org/10.1287/mnsc.1120.1667

Lin, Y., Breugelmans, J., Iversen, M., & Schmidt, D. (2017). An Adaptive Interface Design (AID) for enhanced computer accessibility and rehabilitation. *International Journal of Human-Computer Studies*, *98*, 14–23. https://doi.org/10.1016/j.ijhcs.2016.09.012

Liu, P., Qiu, Y., Hu, J., Tong, J., Zhao, J., & Li, Z. (2020). Expert judgments for performance shaping Factors' multiplier design in human reliability analysis. *Reliability Engineering & System Safety*, *194*, 106343. https://doi.org/10.1016/j.ress.2018.12.022

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of Internal Clustering Validation Measures. *2010 IEEE International Conference on Data Mining*, 911–916. https://doi.org/10.1109/ICDM.2010.35

Liubicich, M. E., Magistro, D., Candela, F., Rabaglietti, E., & Ciairano, S. (2012). Physical Activity, Fine Manual Dexterity and a Coach's Self-Efficacy in a Physical Activity Program for Older Persons Living in Residential Care Facilities. *Psychology*, *3*(5), 384–392. https://doi.org/10.4236/psych.2012.35054

Lyles, C. R., Harris, L. T., Jordan, L., Grothaus, L., Wehnes, L., Reid, R. J., & Ralston, J. D. (2012). Patient Race/Ethnicity and Shared Medical Record Use Among Diabetes Patients. *Medical Care*, *50*(5), 434. https://doi.org/10.1097/MLR.0b013e318249d81b

Ma, J., & Kim, H. M. (2016). Product family architecture design with predictive, data-driven product family design method. *Research in Engineering Design*, *27*(1), 5–21. https://doi.org/10.1007/s00163-015-0201-4

Mahoney, J. M., Kurczewski, N. A., & Froede, E. W. (2015). Design method for multi-user workstations utilizing anthropometry and preference data. *Applied Ergonomics*, *46*, 60–66. https://doi.org/10.1016/j.apergo.2014.07.003

Maier, J. R. A., & Fadel, G. M. (2009). Affordance-based design methods for innovative design, redesign and reverse engineering. *Research in Engineering Design*, *20*(4), 225. https://doi.org/10.1007/s00163-009-0064-7

Manera, K., Hanson, C. S., Gutman, T., & Tong, A. (2019). Consensus Methods: Nominal Group Technique. In P. Liamputtong (Ed.), *Handbook of Research*

Methods in Health Social Sciences (pp. 737–750). Springer. https://doi.org/10.1007/978-981-10-5251-4_100

Mannaru, P., Balasingam, B., Pattipati, K., Sibley, C., & Coyne, J. (2016). Cognitive Context Detection for Adaptive Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 223–227. https://doi.org/10.1177/1541931213601050

Marques, E. A., Baptista, F., Santos, D. A., Silva, A. M., Mota, J., & Sardinha, L. B. (2014). Risk for losing physical independence in older adults: The role of sedentary time, light, and moderate to vigorous physical activity. *Maturitas*, *79*(1), 91–95. https://doi.org/10.1016/j.maturitas.2014.06.012

Marras, W. S. (2012). Basic Biomechanics and Workstation Design. In *Handbook of Human Factors and Ergonomics* (pp. 347–381). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118131350.ch12

MATLAB ver. R2020b (Version R2020b). (2020). [Computer software]. The MathWorks Inc.

Mayberry, L. S., Kripalani, S., Rothman, R. L., & Osborn, C. Y. (2011). Bridging the Digital Divide in Diabetes: Family Support and Implications for Health Literacy. *Diabetes Technology & Therapeutics*, *13*(10), 1005–1012. https://doi.org/10.1089/dia.2011.0055

McCoy, J. G., & Strecker, R. E. (2011). The cognitive cost of sleep lost. *Neurobiology of Learning and Memory*, *96*(4), 564–582. https://doi.org/10.1016/j.nlm.2011.07.004

McDonald, M. (2012). *Market Segmentation: How to Do It and How to Profit from It*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/umdcp/detail.action?docID=1040905

McHenry, J. C., Insel, K. C., Einstein, G. O., Vidrine, A. N., Koerner, K. M., & Morrow, D. G. (2015). Recruitment of Older Adults: Success May Be in the Details. *The Gerontologist*, *55*(5), 845–853. https://doi.org/10.1093/geront/gns079

McLaughlin, A. C., DeLucia, P. R., Drews, F. A., Vaughn-Cooke, M., Kumar, A., Nesbitt, R. R., & Cluff, K. (2020). Evaluating Medical Devices Remotely: Current Methods and Potential Innovations: *Human Factors*. https://doi.org/10.1177/0018720820953644

McParland, D., & Gormley, I. C. (2015). Model Based Clustering for Mixed Data: ClustMD. *ArXiv:1511.01720 [Stat]*. http://arxiv.org/abs/1511.01720

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Meng, A., Nexø, M. A., & Borg, V. (2017). The impact of retirement on age related cognitive decline – a systematic review. *BMC Geriatrics*, *17*(1), 160. https://doi.org/10.1186/s12877-017-0556-7

Menke, A., Casagrande, S., Geiss, L., & Cowie, C. C. (2015). Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA*, *314*(10), 1021–1029. https://doi.org/10.1001/jama.2015.10029

Mesgari, M., Okoli, C., & Guinea, A. O. de. (2015, June 26). Affordance-based User Personas: A mixed-method Approach to Persona Development. *AMCIS 2015 Proceedings*. https://aisel.aisnet.org/amcis2015/HCI/GeneralPresentations/1

Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, *32*(5), 417–430. https://doi.org/10.1016/j.destud.2011.03.003

Miaskiewicz, T., Sumner, T., & Kozar, K. A. (2008). A latent semantic analysis methodology for the identification and creation of personas. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1501–1510. https://doi.org/10.1145/1357054.1357290

Michalek, J. J., Ceryan, O., Papalambros, P. Y., & Koren, Y. (2005). Balancing Marketing and Manufacturing Objectives in Product Line Design. *Journal of Mechanical Design*, *128*(6), 1196–1204. https://doi.org/10.1115/1.2336252

Miranda, A. T. (2017). Using Conditional Probabilities to Understand "Human Error" in Military Aviation Mishaps. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1619–1620. https://doi.org/10.1177/1541931213601889

Miyake, Y., Kobayashi, R., Kelepecz, D., & Nakajima, M. (2013). Core exercises elevate trunk stability to facilitate skilled motor behavior of the upper extremities. *Journal of Bodywork and Movement Therapies*, *17*(2), 259–265. https://doi.org/10.1016/j.jbmt.2012.06.003

Mokhtarian, H., Coatanéa, E., & Paris, H. (2017). Function modeling combined with physics-based reasoning for assessing design options and supporting innovative ideation. *AI EDAM*, *31*(4), 476–500. https://doi.org/10.1017/S0890060417000403

Morelli, F., Neugebauer, J. M., Haynes, C. A., Fry, T. C., Ortega, S. V., Struve, D. J., LaFiandra, M. E., & Larkin, G. B. (2017). Shooter–System Performance Variability as a Function of Recoil Dynamics. *Human Factors*, *59*(6), 973–985. https://doi.org/10.1177/0018720817700537

Morelli, L., Guadagni, S., Franco, G. D., Palmeri, M., Candio, G. D., & Mosca, F. (2016). Da Vinci single site© surgical platform in clinical practice: A systematic review. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *12*(4), 724–734. https://doi.org/10.1002/rcs.1713

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, *111*(20), 7176–7184. https://doi.org/10.1073/pnas.1319946111

Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: An overview, II. *WIREs Data Mining and Knowledge Discovery*, *7*(6), e1219. https://doi.org/10.1002/widm.1219

Nadadur, G., Raschke, U., & Parkinson, M. B. (2016). A quantile-based anthropometry synthesis technique for global user populations. *International Journal of Industrial Ergonomics*, *53*(Supplement C), 167–178. https://doi.org/10.1016/j.ergon.2015.11.009

Nagel, R. L., Bohm, M. R., Linsey, J. S., & Riggs, M. K. (2015). Improving Students' Functional Modeling Skills: A Modeling Approach and a Scoring Rubric.

Journal of Mechanical Design, *137*(5), 051102-051102–051113. https://doi.org/10.1115/1.4029585

National Aeronautics and Space Administration. (2014). *Human Integration Design Handbook (NASA/SP-2010-3407)*. https://www.nasa.gov/feature/human-integration-design

National Aeronautics and Space Administration. (2019). *Spaceflight Human-System Standard Volume 2: Human Factors, Habitability, and Environmental Health (NASA-STD-3001)*. https://www.nasa.gov/hhp/standards

Neate, T., Bourazeri, A., Roper, A., Stumpf, S., & Wilson, S. (2019). Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300880

Nelson, L. A., Coston, T. D., Cherrington, A. L., & Osborn, C. Y. (2016). Patterns of user engagement with mobile- and web-delivered self-care interventions for adults with T2DM: A review of the literature. *Current Diabetes Reports*, *16*(7), 66. https://doi.org/10.1007/s11892-016-0755-1

Ng, F. L., Saxena, M., Mahfoud, F., Pathak, A., & Lobo, M. D. (2016). Device-based Therapy for Hypertension. *Current Hypertension Reports*, *18*. https://doi.org/10.1007/s11906-016-0670-5

Niazi, B., Khusro, S., Khan, A., & Alam, I. (2016). A Touch Sensitive Keypad Layout for Improved Usability of Smartphones for the Blind and Visually Impaired Persons. In R. Silhavy, R. Senkerik, Z. K. Oplatkova, P. Silhavy, & Z. Prokopova (Eds.), *Artificial Intelligence Perspectives in Intelligent Systems* (pp. 427–436). Springer International Publishing. https://doi.org/10.1007/978-3-319-33625-1_38

Oh, S., Haberl, J. S., & Baltazar, J.-C. (2020). Analysis methods for characterizing energy saving opportunities from home automation devices using smart meter data. *Energy and Buildings*, *216*, 109955. https://doi.org/10.1016/j.enbuild.2020.109955

O'Hagan, A. (2019). Expert Knowledge Elicitation: Subjective but Scientific. *The American Statistician*, *73*(sup1), 69–81. https://doi.org/10.1080/00031305.2018.1518265

OHara, J., Higgins, J. C., Fleger, S. A., & Pieringer, P. A. (2012). *Human Factors Engineering Program Review Model (NUREG-0711)* (p. 147) [Standard]. Office of Nuclear Regulatory Research. https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0711/

Otto, K., Hölttä-Otto, K., Simpson, T. W., Krause, D., Ripperda, S., & Ki Moon, S. (2016). Global Views on Modular Design Research: Linking Alternative Methods to Support Modular Product Family Concept Development. *Journal of Mechanical Design*, *138*(7), 071101-071101–071116. https://doi.org/10.1115/1.4033654

Ozcelik, D., Quevedo-Fernandez, J., Thalen, J., & Terken, J. (2011). Engaging users in the early phases of the design process: Attitudes, concerns and challenges from industrial practice. *Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces*, 1–8. https://doi.org/10.1145/2347504.2347519

Paço, A. M. F. do, & Raposo, M. L. B. (2010). Green consumer market segmentation: Empirical findings from Portugal. *International Journal of Consumer Studies*, *34*(4), 429–436. https://doi.org/10.1111/j.1470-6431.2010.00869.x

Pahl, G., Beitz, W., Feldhusen, J., & Grote, K.-H. (2007). *Engineering Design: A Systematic Approach* (3rd ed.). Springer-Verlag. https://www.springer.com/gp/book/9781846283185

Pallant, J., Sands, S., & Karpen, I. (2020). Product customization: A profile of consumer demand. *Journal of Retailing and Consumer Services*, *54*, 102030. https://doi.org/10.1016/j.jretconser.2019.102030

Pandya, D., Podofillini, L., Emert, F., Lomax, A. J., Dang, V. N., & Sansavini, G. (2020). Quantification of a human reliability analysis method for radiotherapy applications based on expert judgment aggregation. *Reliability Engineering & System Safety*, *194*, 106489. https://doi.org/10.1016/j.ress.2019.05.001

Parkinson, M. B., & Reed, M. P. (2010). Creating virtual user populations by analysis of anthropometric data. *International Journal of Industrial Ergonomics*, *40*(1), 106–111. https://doi.org/10.1016/j.ergon.2009.07.003

Patel, A., Kramer, W. S., Flynn, M., Summers, J. D., & Shuffler, M. L. (2020). Function Modeling: A Modeling Behavior Analysis of Pause Patterns. *Journal of Mechanical Design*, *142*(111402). https://doi.org/10.1115/1.4046999

Pérez-Jover, V., Sala-González, M., Guilabert, M., & Mira, J. J. (2019). Mobile Apps for Increasing Treatment Adherence: Systematic Review. *Journal of Medical Internet Research*, *21*(6), e12505. https://doi.org/10.2196/12505

Phillips, A. W., Smith, S. G., & Straus, C. M. (2013). Driving Deeper Learning by Assessment: An Adaptation of the Revised Bloom's Taxonomy for Medical Imaging in Gross Anatomy. *Academic Radiology*, *20*(6), 784–789. https://doi.org/10.1016/j.acra.2013.02.001

Pirmoradi, Z., Hajikolaei, K. H., & Wang, G. G. (2015). Designing scalable product families by the radial basis function–high-dimensional model representation metamodelling technique. *Engineering Optimization*, *47*(10), 1423–1439. https://doi.org/10.1080/0305215X.2014.971776

Pirmoradi, Z., Wang, G., & Simpson, T. (2014). A Review of Recent Literature in Product Family Design and Platform-Based Product Development. In *Product Platform and Product Family Design: Methods and Applications* (pp. 1–46). https://doi.org/10.1007/978-1-4614-7937-6_1

Pliner, E. M., Sturnieks, D. L., Beschorner, K. E., Redfern, M. S., & Lord, S. R. (2021). Individual factors that influence task performance on a stepladder in older people. *Safety Science*, *136*, 105152. https://doi.org/10.1016/j.ssci.2020.105152

Podani, J. (1999). Extending Gower's General Coefficient of Similarity to Ordinal Characters. *Taxon*, *48*(2), 331–340. JSTOR. https://doi.org/10.2307/1224438

Poirson, E., & Parkinson, M. (2014). Estimated anthropometry for male commercial pilots in Europe and an approach to its use in seat design. *International Journal of Industrial Ergonomics*, *44*(5), 769–776. https://doi.org/10.1016/j.ergon.2014.05.003

Pomarici, E., Lerro, M., Chrysochou, P., Vecchio, R., & Krystallis, A. (2017). One size does (obviously not) fit all: Using product attributes for wine market segmentation. *Wine Economics and Policy*, *6*(2), 98–106. https://doi.org/10.1016/j.wep.2017.09.001

Präg, P., Mills, M. C., & Wittek, R. (2016). Subjective socioeconomic status and health in cross-national comparison. *Social Science & Medicine*, *149*, 84–92. https://doi.org/10.1016/j.socscimed.2015.11.044

Princi, E., & Krämer, N. C. (2020). Out of Control – Privacy Calculus and the Effect of Perceived Control and Moral Considerations on the Usage of IoT Healthcare Devices. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.582054

Pritchard, J. N., & Nicholls, C. (2014). Emerging Technologies for Electronic Monitoring of Adherence, Inhaler Competence, and True Adherence. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, *28*(2), 69–81. https://doi.org/10.1089/jamp.2014.1163

Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2014a). Measuring Human-Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*, *8*(1), 52–77. https://doi.org/10.1177/1555343413490166

Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2014b). Modeling Human–Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*, *8*(1), 33–51. https://doi.org/10.1177/1555343413490944

Privitera, M. B. (2020). Do we really know our users? A scientific approach to defining user groups for human factors studies. *Proceedings of the*

International Symposium on Human Factors and Ergonomics in Health Care, *9*, 139–143. https://doi.org/10.1177/2327857920091038

Proud, J. K., Lai, D. T. H., Mudie, K. L., Carstairs, G. L., Billing, D. C., Garofolini, A., & Begg, R. K. (2020). Exoskeleton Application to Military Manual Handling Tasks. *Human Factors*, 0018720820957467. https://doi.org/10.1177/0018720820957467

Proud, R. W., Hart, J. J., & Mrozinski, R. B. (2003). *Methods for Determining the Level of Autonomy to Design into a Human Spaceflight Vehicle: A Function Specific Approach*. National Aeronautics and Space Administration. https://apps.dtic.mil/docs/citations/ADA515467

Qayyum, A., Qadir, J., Bilal, M., & Al-Fuqaha, A. (2020). Secure and Robust Machine Learning for Healthcare: A Survey. *ArXiv:2001.08103 [Cs, Eess, Stat]*. http://arxiv.org/abs/2001.08103

Qualtrics XM (September 2020). (2020). [Computer software]. Qualtrics.

Radwin, R. G., Chourasia, A. O., Howery, R. S., Fronczak, F. J., Yen, T. Y., Subedi, Y., & Sesto, M. E. (2014). Factors Influencing Power Hand Tool Fastening Accuracy and Reaction Forces. *Human Factors*, *56*(4), 657–668. https://doi.org/10.1177/0018720813507952

Ramachandran, R., Caldwell, B. W., & Mocko, G. M. (2011). *A User Study to Evaluate the Function Model and Function Interaction Model for Concept Generation*. 273–284. https://doi.org/10.1115/DETC2011-47660

Ramasubbareddy, S., Srinivas, T. A. S., Govinda, K., & Manivannan, S. S. (2020). Comparative Study of Clustering Techniques in Market Segmentation. In H.

S. Saini, R. Sayal, R. Buyya, & G. Aliseri (Eds.), *Innovations in Computer Science and Engineering: Proceedings of 7th ICICSE* (pp. 117–125). Springer. https://doi.org/10.1007/978-981-15-2043-3_15

Razavi, R. (2020). Personality segmentation of users through mining their mobile usage patterns. *International Journal of Human-Computer Studies*, *143*, 102470. https://doi.org/10.1016/j.ijhcs.2020.102470

Reddy, K., Byrne, D., Breen, D., Lydon, S., & O'Connor, P. (2020). The application of human reliability analysis to three critical care procedures. *Reliability Engineering & System Safety*, *203*, 107116. https://doi.org/10.1016/j.ress.2020.107116

Rejeski, W. J., Marsh, A. P., Chmelo, E., & Rejeski, J. J. (2010). Obesity, intentional weight loss and physical disability in older adults. *Obesity Reviews*, *11*(9), 671–685. https://doi.org/10.1111/j.1467-789X.2009.00679.x

Reyna, V. F. (2008). A Theory of Medical Decision Making and Health: Fuzzy Trace Theory. *Medical Decision Making*, *28*(6), 850–865. https://doi.org/10.1177/0272989X08327066

Roberts, N. (1976). Further Verification of Bloom's Taxonomy. *The Journal of Experimental Education*, *45*(1), 16–19. https://doi.org/10.1080/00220973.1976.11011556

Roma, M. S. G., & de Vilhena Garcia, E. (2020). Medical device usability: Literature review, current status, and challenges. *Research on Biomedical Engineering*, *36*(2), 163–170. https://doi.org/10.1007/s42600-019-00037-8

Roman, L. C., Ancker, J. S., Johnson, S. B., & Senathirajah, Y. (2017). Navigation in the electronic health record: A review of the safety and usability literature. *Journal of Biomedical Informatics*, *67*, 69–79. https://doi.org/10.1016/j.jbi.2017.01.005

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research*, *2*(4), 320–332. https://doi.org/10.1504/IJBHR.2011.043414

Rossiter, A. (2020). The impact of robotics and autonomous systems (RAS) across the conflict spectrum. *Small Wars & Insurgencies*, *31*(4), 691–700. https://doi.org/10.1080/09592318.2020.1743481

Rudman, D. L., Gold, D., McGrath, C., Zuvela, B., Spafford, M. M., & Renwick, R. (2016). "Why would I want to go out?": Age-related Vision Loss and Social Participation. *Canadian Journal on Aging / La Revue Canadienne Du Vieillissement*, *35*(4), 465–478. https://doi.org/10.1017/S0714980816000490

Saidi, T., Mutswangwa, C. T., & Douglas, T. S. (2019). Design Thinking as a Complement to Human Factors Engineering for Enhancing Medical Device Usability. *Engineering Studies*, *11*(1), 34–50. https://doi.org/10.1080/19378629.2019.1567521

Salminen, J., Guan, K., Jung, S.-G., Chowdhury, S. A., & Jansen, B. J. (2020). A Literature Review of Quantitative Persona Creation. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376502

Salminen, J., Santos, J. M., Kwak, H., An, J., Jung, S., & Jansen, B. J. (2020). Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas. *International Journal of Human-Computer Studies*, *141*, 102437. https://doi.org/10.1016/j.ijhcs.2020.102437

Sangelkar, S., Cowen, N., & McAdams, D. (2012). User activity – product function association based design rules for universal products. *Design Studies*, *33*(1), 85–110. https://doi.org/10.1016/j.destud.2011.06.002

Santoso, L. W. (2021). Adaptive Educational Resources Framework for ELearning Using Rule-Based System. In T. Senjyu, P. N. Mahalle, T. Perumal, & A. Joshi (Eds.), *Information and Communication Technology for Intelligent Systems* (pp. 385–396). Springer. https://doi.org/10.1007/978-981-15-7078-0_36

Scheuringer, M., Stucki, G., Huber, E. O., Brach, M., Schwarzkopf, S. R., Kostanjsek, N., & Stoll, T. (2005). ICF Core Set for patients with musculoskeletal conditions in early post-acute rehabilitation facilities. *Disability and Rehabilitation*, *27*(7–8), 405–410. https://doi.org/10.1080/09638280400014006

Schubert, E., & Rousseeuw, P. J. (2019). Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. *ArXiv:1810.05691 [Cs, Stat]*, *11807*, 171–187. https://doi.org/10.1007/978-3-030-32047-8_16

Seepersad, C. C., Pedersen, K., Emblemsvåg, J., Bailey, R., Allen, J. K., & Mistree, F. (2006). *The Validation Square: How Does One Verify and Validate a Design Method?* https://doi.org/10.1115/1.802469.ch25

Seidler, R. D., Bernard, J. A., Burutolu, T. B., Fling, B. W., Gordon, M. T., Gwin, J. T., Kwak, Y., & Lipps, D. B. (2010). Motor control and aging: Links to age-related brain structural, functional, and biochemical effects. *Neuroscience & Biobehavioral Reviews*, *34*(5), 721–733. https://doi.org/10.1016/j.neubiorev.2009.10.005

Selb, M., Escorpizo, R., Kostanjsek, N., Stucki, G., Üstün, B., & Cieza, A. (2015). A guide on how to develop an International Classification of Functioning, Disability and Health Core Set. *European Journal of Physical and Rehabilitation Medicine*, *51*(1), 105–117.

Senefeld, J., Yoon, T., & Hunter, S. K. (2017). Age differences in dynamic fatigability and variability of arm and leg muscles: Associations with physical function. *Experimental Gerontology*, *87*, 74–83. https://doi.org/10.1016/j.exger.2016.10.008

Shan, R., Sarkar, S., & Martin, S. S. (2019). Digital health technology and mobile devices for the management of diabetes mellitus: State of the art. *Diabetologia*, *62*(6), 877–887. https://doi.org/10.1007/s00125-019-4864-7

Sheridan, T. B. (2000). Function allocation: Algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies*, *52*(2), 203–216. https://doi.org/10.1006/ijhc.1999.0285

Sheridan, T. B. (2011). Adaptive Automation, Level of Automation, Allocation Authority, Supervisory Control, and Adaptive Control: Distinctions and Modes of Adaptation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *41*(4), 662–667. https://doi.org/10.1109/TSMCA.2010.2093888

Sherkat, M., Miller, T., & Mendoza, A. (2016). Does it Fit Me Better? User Segmentation in Requirements Engineering. *2016 23rd Asia-Pacific Software Engineering Conference (APSEC)*, 65–72. https://doi.org/10.1109/APSEC.2016.020

Shoaib, M., Hussain, I., & Mirza, H. T. (2020). Automatic switching between speech and non-speech: Adaptive auditory feedback in desktop assistance for the visually impaired. *Universal Access in the Information Society*, *19*(4), 813–823. https://doi.org/10.1007/s10209-019-00696-5

Showell, C. (2017). Barriers to the use of personal health records by patients: A structured review. *PeerJ*, *5*, e3268. https://doi.org/10.7717/peerj.3268

Siebra, C., Gouveia, T. B., Macedo, J., da Silva, F. Q. B., Santos, A. L. M., Correia, W., Penha, M., Anjos, M., & Florentin, F. (2017). Toward accessibility with usability: Understanding the requirements of impaired uses in the mobile context. *Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication*, 1–8. https://doi.org/10.1145/3022227.3022233

Siebra, C., Gouveia, T., Macedo, J., Correia, W., Penha, M., Silva, F., Santos, A., Anjos, M., & Florentin, F. (2015). Usability requirements for mobile accessibility: A study on the vision impairment. *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, 384–389. https://doi.org/10.1145/2836041.2841213

Silva, C., Masci, P., Zhang, Y., Jones, P., & Campos, J. C. (2019). A use error taxonomy for improving human-machine interface design in medical devices. *ACM SIGBED Review*, *16*(2), 24–30. https://doi.org/10.1145/3357495.3357498

Simpson, T. W., Bobuk, A., Slingerland, L. A., Brennan, S., Logan, D., & Reichard, K. (2012). From user requirements to commonality specifications: An integrated approach to product family design. *Research in Engineering Design*, *23*(2), 141–153. https://doi.org/10.1007/s00163-011-0119-4

Simpson, T. W., Maier, J. R., & Mistree, F. (2001). Product platform design: Method and application. *Research in Engineering Design*, *13*(1), 2–22. https://doi.org/10.1007/s001630100002

Sinha, K., & Suh, E. S. (2018). Pareto-optimization of complex system architecture for structural complexity and modularity. *Research in Engineering Design*, *29*(1), 123–141. https://doi.org/10.1007/s00163-017-0260-9

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi Method for Graduate Research. *Journal of Information Technology Education: Research*, *6*(1), 1–21.

Sobol, N. A., Hoffmann, K., Vogel, A., Lolk, A., Gottrup, H., Høgh, P., Hasselbalch, S. G., & Beyer, N. (2016). Associations between physical function, dual-task performance and cognition in patients with mild Alzheimer's disease. *Aging*

& Mental Health, *20*(11), 1139–1146. https://doi.org/10.1080/13607863.2015.1063108

Soozandehfar, S. M. A., & Adeli, M. R. (2016). A Critical Appraisal of Bloom's Taxonomy. *American Research Journal of English and Literature*, *2*, 9.

Soria Zurita, N. F., Stone, R. B., Onan Demirel, H., & Tumer, I. Y. (2020). Identification of Human–System Interaction Errors During Early Design Stages Using a Functional Basis Framework. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, *6*(1). https://doi.org/10.1115/1.4044787

Srivastava, A. N., & Schumann, J. (2013). Software health management: A necessity for safety critical systems. *Innovations in Systems and Software Engineering*, *9*(4), 219–233. https://doi.org/10.1007/s11334-013-0212-0

Stevenson, P. D., Mattson, C. A., Bryden, K. M., & MacCarty, N. A. (2018). Toward a Universal Social Impact Metric for Engineered Products That Alleviate Poverty. *Journal of Mechanical Design*, *140*(4). https://doi.org/10.1115/1.4038925

Stone, R., & Wood, K. L. (2000). Development of a Functional Basis for Design. *Journal of Mechanical Design*, *122*(4), 359–370. https://doi.org/10.1115/1.1289637

Stucki, A., Daansen, P., Fuessl, M., Cieza, A., Huber, E., Atkinson, R., Kostanjsek, N., Stucki, G., & Ruof, J. (2004). ICF Core Sets for obesity. *Journal of Rehabilitation Medicine*, *44 Suppl*, 107–113. https://doi.org/10.1080/16501960410016064

Stucki, A., Stoll, T., Cieza, A., Weigl, M., Giardini, A., Wever, D., Kostanjsek, N., & Stucki, G. (2004). ICF Core Sets for obstructive pulmonary diseases. *Journal of Rehabilitation Medicine*, *44 Suppl*, 114–120. https://doi.org/10.1080/16501960410016794

Sun, X., Houssin, R., Renaud, J., & Gardoni, M. (2018). Towards a human factors and ergonomics integration framework in the early product design phase: Function-Task-Behaviour. *International Journal of Production Research*, *56*(14), 4941–4953. https://doi.org/10.1080/00207543.2018.1437287

Tanenbaum, M. L., Adams, R. N., Iturralde, E., Hanes, S. J., Barley, R. C., Naranjo, D., & Hood, K. K. (2018). From Wary Wearers to d-Embracers: Personas of Readiness to Use Diabetes Devices. *Journal of Diabetes Science and Technology*, *12*(6), 1101–1107. https://doi.org/10.1177/1932296818793756

Tipton, E., & Matlen, B. J. (2019). Improved Generalizability Through Improved Recruitment: Lessons Learned From a Large-Scale Randomized Trial. *American Journal of Evaluation*, *40*(3), 414–430. https://doi.org/10.1177/1098214018810519

Tomiyama, T., Beek, T. J. van, Cabrera, A. A. A., Komoto, H., D&#39, V., & Amelio. (2013). Making function modeling practically usable. *AI EDAM*, *27*(3), 301–309. https://doi.org/10.1017/S0890060413000309

Tomlinson, S. M. (2016). Perceptions of accessibility and usability by blind or visually impaired persons: A pilot study. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–4. https://doi.org/10.1002/pra2.2016.14505301120

Toplak, M. E., Sorge, G. B., Benoit, A., West, R. F., & Stanovich, K. E. (2010). Decision-making and cognitive abilities: A review of associations between Iowa Gambling Task performance, executive functions, and intelligence. *Clinical Psychology Review*, *30*(5), 562–581. https://doi.org/10.1016/j.cpr.2010.04.002

Trief, P. M., Izquierdo, R., Eimicke, J. P., Teresi, J. A., Goland, R., Palmas, W., Shea, S., & Weinstock, R. S. (2013). Adherence to diabetes self care for white, African-American and Hispanic American telemedicine participants: 5 year results from the IDEATel project. *Ethnicity & Health*, *18*(1), 83–96. https://doi.org/10.1080/13557858.2012.700915

Tucker, C. S., Kim, H. M., Barker, D. E., & Zhang, Y. (2010). A ReliefF attribute weighting and X-means clustering methodology for top-down product family optimization. *Engineering Optimization*, *42*(7), 593–616. https://doi.org/10.1080/03052150903353328

Umeda, Y., Ishii, M., Yoshioka, M., Shimomura, Y., & Tomiyama, T. (1996). Supporting conceptual design based on the function-behavior-state modeler. *AI EDAM*, *10*(4), 275–288. https://doi.org/10.1017/S0890060400001621

Unal, M., Warn, G. P., & Simpson, T. W. (2017). Quantifying the Shape of Pareto Fronts During Multi-Objective Trade Space Exploration. *Journal of Mechanical Design*, *140*(021402). https://doi.org/10.1115/1.4038005

UyBico, S. J., Pavel, S., & Gross, C. P. (2007). Recruiting Vulnerable Populations into Research: A Systematic Review of Recruitment Interventions. *Journal of*

General Internal Medicine, *22*(6), 852–863. https://doi.org/10.1007/s11606-007-0126-3

Van Acker, B. B., Bombeke, K., Durnez, W., Parmentier, D. D., Mateus, J. C., Biondi, A., Saldien, J., & Vlerick, P. (2020). Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure. *International Journal of Industrial Ergonomics*, *75*, 102891. https://doi.org/10.1016/j.ergon.2019.102891

Verenna, A.-M. A., Noble, K. A., Pearson, H. E., & Miller, S. M. (2018). Role of comprehension on performance at higher levels of Bloom’s taxonomy: Findings from assessments of healthcare professional students. *Anatomical Sciences Education*, *11*(5), 433–444. https://doi.org/10.1002/ase.1768

Viana, G., & Robert, J.-M. (2016). The Practitioners’ Points of View on the Creation and Use of Personas for User Interface Design. In M. Kurosu (Ed.), *Human-Computer Interaction. Theory, Design, Development and Practice* (pp. 233–244). Springer International Publishing. https://doi.org/10.1007/978-3-319-39510-4_22

Wang, L., Li, L., Cai, H., Xu, L., Xu, B., & Jiang, L. (2018). Analysis of Regional Group Health Persona Based on Image Recognition. *2018 Sixth International Conference on Enterprise Systems (ES)*, 166–171. https://doi.org/10.1109/ES.2018.00033

Wang, N., Xu, C., Du, X., & Zhang, M. (2018). A risk assessment method of deep excavation based on Bayesian analysis and expert elicitation. *International*

Journal of System Assurance Engineering and Management, *9*(2), 452–466. https://doi.org/10.1007/s13198-017-0689-2

Wang, Q., Tang, D., Li, S., Yang, J., Salido, M. A., Giret, A., & Zhu, H. (2019). An Optimization Approach for the Coordinated Low-Carbon Design of Product Family and Remanufactured Products. *Sustainability*, *11*(2), 460. https://doi.org/10.3390/su11020460

Wäppling, D., Feng, X., Andersson, H., Pettersson, M., Lunden, B., & Weström, J. (2011). *Simultaneous Requirement and Design Optimization of an Industrial Robot Family Using Multi-Objective Optimization*. 815–823. https://doi.org/10.1115/DETC2011-47819

Warin, B., Kolski, C., & Toffolon, C. (2018). Living persona technique applied to HCI education. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 51–59. https://doi.org/10.1109/EDUCON.2018.8363208

Wedel, M., & Kamakura, W. A. (2012). *Market Segmentation: Conceptual and Methodological Foundations* (Softcover reprint of the original 2nd ed. 2000 edition). Springer.

Weppner, W. G., Ralston, J. D., Koepsell, T. D., Grothaus, L. C., Reid, R. J., Jordan, L., & Larson, E. B. (2010). Use of a Shared Medical Record With Secure Messaging by Older Patients With Diabetes. *Diabetes Care*, *33*(11), 2314–2319. https://doi.org/10.2337/dc10-1124

Whitney, P., Hinson, J. M., Jackson, M. L., & Van Dongen, H. P. A. (2015). Feedback Blunting: Total Sleep Deprivation Impairs Decision Making that

Requires Updating Based on Feedback. *Sleep*, *38*(5), 745–754. https://doi.org/10.5665/sleep.4668

Wildenbos, G. A., Peute, L., & Jaspers, M. (2018). Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). *International Journal of Medical Informatics*, *114*, 66–75. https://doi.org/10.1016/j.ijmedinf.2018.03.012

Wildner, M., Quittan, M., Portenier, L., Wilke, S., Boldt, C., Stucki, G., Kostanjsek, N., & Grill, E. (2005). ICF Core Set for patients with cardiopulmonary conditions in early post-acute rehabilitation facilities. *Disability and Rehabilitation*, *27*(7–8), 397–404. https://doi.org/10.1080/09638280400013958

World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. Geneva: World Health Organization.

Wu, J., & Azarm, S. (2000). Metrics for Quality Assessment of a Multiobjective Design Optimization Solution Set. *Journal of Mechanical Design*, *123*(1), 18–25. https://doi.org/10.1115/1.1329875

Wu, P., O'Malley, J. T., Gruttola, V. de, & Liberman, M. C. (2020). Age-Related Hearing Loss Is Dominated by Damage to Inner Ear Sensory Cells, Not the Cellular Battery That Powers Them. *Journal of Neuroscience, 40*(33), 6357–6366. https://doi.org/10.1523/JNEUROSCI.0937-20.2020

Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science, 2*(2), 165–193. https://doi.org/10.1007/s40745-015-0040-1

Xu, Y., & Huijun, Z. (2007). Function principles for a mechatronic system design. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *221*(6), 1065–1076. https://doi.org/10.1243/09544054JEM649

Yannou, B., Yvars, P.-A., Hoyle, C., & Chen, W. (2013). Set-based design by simulation of usage scenario coverage. *Journal of Engineering Design*, *24*(8), 575–603. https://doi.org/10.1080/09544828.2013.780201

Yildirim, U., Campean, F., & Williams, H. (2017). Function modeling using the system state flow diagram. *AI EDAM*, *31*(4), 413–435. https://doi.org/10.1017/S0890060417000294

Zhang, G., & Thai, V. V. (2016). Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review. *Safety Science*, *87*, 53–62. https://doi.org/10.1016/j.ssci.2016.03.019

Zhao, S., Pan, G., Tao, J., Luo, Z., Li, S., & Wu, Z. (2020). Understanding Smartphone Users From Installed App Lists Using Boolean Matrix Factorization. *IEEE Transactions on Cybernetics*, 1–14. https://doi.org/10.1109/TCYB.2020.2967644

Zhou, Y., & Liu, L. (2020). Development and Testing of a User-adaptive Ankle Foot Orthosis. *2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM)*, 582–587. https://doi.org/10.1109/ICARM49381.2020.9195386

Zimmer, J. C., Tams, S., Craig, K., Thatcher, J., & Pak, R. (2015). The role of user age in task performance: Examining curvilinear and interaction effects of user age, expertise, and interface design on mistake making. *Journal of Business Economics*, *85*(4), 323–348. https://doi.org/10.1007/s11573-015-0762-8