Quantitative Evaluation of Personas as Information

Christopher N. Chapman[1], Edwin Love[2], Russell P. Milham[1], Paul ElRif[1], James L. Alford[3]

[1] *1 Microsoft Way, Redmond, WA 98052,* chris.chapman|russ.milham|pelrif@microsoft.com
[2] *270 Mackenzie, Box 353200, Univ. of Washington, Seattle, WA 98195,* edwinl@u.washington.edu,
[3] *Volt Information Sciences, 11261 Willows Rd., NE Ste. 200, Redmond, WA 98052,* jalford0974@yahoo.com

The personas method is said to present information about people of interest for product design. We propose a formal model to understand persona information in terms of factual attributes. Using an analytic model, we show that the expected prevalence rate of persona descriptions decreases rapidly as attributes are added. We then evaluate this expectation empirically. Using six survey datasets ranging from N=268 to N=10307 respondents and two simulated datasets, we determine the prevalence rates of 10000 randomly generated persona-like descriptions per dataset. Consistent with prediction, we observe decreasing prevalence rates as attributes are added. Pearson's r for observed vs. predicted prevalence, transformed to multinormality, ranges $r(9998)=0.394$ to $r(9998)=0.869$ in the sampled datasets (all $p < 0.001$). Because descriptions with many attributes are likely to represent few people, we suggest that personas should be assessed empirically before they are assumed to describe real groups of people.

PERSONAS AS INFORMATION

The personas method is widely used by technology designers and human-computer interaction researchers to describe users and customers (Cooper, 1999; Courage & Baxter, 2005; Holtzblatt, Wendell & Wood, 2006; Mulder & Yaar, 2006; Pruitt & Adlin, 2006). A persona consists of a description of a fictional person who represents an important customer or user group for the product, and typically presents information about demographics, behavior, product usage, and product-related goals, tasks, attitudes, etc.

Originally, personas were developed primarily as a tool for design inspiration (Cooper, 1999), but today they are more commonly intended and understood to present information about real user groups of importance. Cooper, Reimann, and Cronin (2007) state that "Personas represent groups of users" and "Personas must be based on research" (p. 82). Pruitt and Grudin (2003) regard personas as "a conduit for information about users and work settings derived from ethnographies, market research, usability studies, interviews, observations, and so on." (p. 10).

Understanding personas as a conduit for information poses many questions. Two fundamental questions involve how to understand the kind of information that a persona presents, and how to determine how many people are represented by a description that combines multiple attributes. The present work explores these issues by examining three basic questions. Is there a formal way to understand the kind of factual information conveyed by a persona? Given a formal understanding, is it possible to estimate the number of users that a persona might represent? In real datasets, how many people are represented by potential persona descriptions?

We propose a formal model for understanding the information in a persona such that it can, to some extent, be mapped to empirical data. We call such specific sets of information "persona-like descriptions" and examine the predicted general prevalence of such descriptions. Finally, we examine the actual prevalence of persona-like descriptions in six real and two simulated datasets.

A PROPOSED FORMAL MODEL

To explore personas in a systematic and empirical fashion, one must determine whether personas have a common structure. Chapman and Milham (2006) suggest that the informational content of a persona comprises – in part – a conjunction of asserted facts, or what we call *attributes*. Following is an excerpt from a published persona:

> Kathleen is 33yrs old and lives in Seattle. She's a stay-at-home mom with two children: Katie, 7, and Andrew, 4. She drives the kids to school (usually carpooling with 2-3 other kids) in her Volvo wagon. Kathleen is thinking about buying [a] rear-seat entertainment system… (Brechin, 2002, p. 1)

The informational content here can be represented as a list of asserted attributes, such as:

[persona] is named Kathleen, and
[persona] is 33 years old, and

[persona] lives in Seattle, and …

In practice, we do not evaluate open-ended point values, e.g., "33 years old," because they match few people. As described below, we represent values as nominal or ordinal ranges (for instance, "31-40 years old"). However, we discuss point values in this section to illustrate the cited example.

Such a list of attributes shows the structure of a conjunction of independent assertions:

A1 ("Kathleen") & B1 ("33 years old") & …

If this represents a claim about real user groups, as suggested by Cooper and others, then it can be assessed empirically, and one can determine how many people are selected from a population by such a conjunction of assertions. Logically, this takes the form of a quantifier selecting people:

[persona] = $x \mid A1(x) \wedge B1(x) \wedge C1(x) \wedge \ldots$

That is, the persona group consists of all people such that A1 is true of each person, and B1 is true of each, and C1 is true of each, and so forth. The question of empirical prevalence, then, is how many people match this logical expression.

Such information could represent part of a given persona. Along with attributes, there could be other content such as photographs that do not map well to logical expressions. Our model applies to the subset of information content in a persona that consists of specific assertions.

THE STRUCTURE OF PERSONA ATTRIBUTES

From the point of view of empirical validation, we propose that the best way to consider individual persona attributes is to view them as nominal categorical variables. A descriptor such as "33 years old" is an example of a nominal (or ordinal) variable that might be expressed in a more general way as "31-40 years old." Someone matches this individual attribute when his or her age falls within the same range as the persona's stated age.

A brief formal description of such a *single* attribute *Ax* can be given as follows:

1. $x \in \{[people\ of\ interest]\}$
   x is a person in the population of interest.

2. $Ax \in \{A1, A2, A3, \ldots An\}$
   A is a set of attributes about x, from an enumerated list of possible attribute levels.

3. $\forall x\ (A1(x)\ \vee A2(x)\ \vee A3(x)\ \vee \ldots An(x))$
   At least one attribute/level in that list is true.

4. $\forall Ay \in \{A1 \ldots An\}\ \forall Az \in \{A1 \ldots An\}, Ay \neq Az \rightarrow \Pr(Ay\ \wedge\ Az) =\ 0$
   Attribute levels are mutually exclusive.

A common structure that meets these formal requirements is a nominal survey item response. A nominal survey item is answered by a person (meeting condition 1 above); it has a predetermined or observed list of possible values (meeting condition 2); respondents choose an answer from the list (satisfying condition 3); and they choose only one answer (satisfying condition 4). Nothing precludes attributes from being ordinal, interval, or ratio items. However, they must be *at least* nominal.

A potential empirically assessable translation of a persona, then, is a list of survey items and responses, such that the persona is defined (in part) by the combined list of items and single responses to each. For example, the persona known as "Kathleen" might be translated as matching nominal answers of "31-40 years old", "Living in the US Pacific Northwest" and so forth on a survey. This allows us to link a persona to an empirical set of data, when at least some of the persona's attributes are reflected by items in a survey. We call such a set of jointly specified attributes, such as might appear in a written persona, a *persona-like description*.

EXPLORING THE PREVALENCE OF PERSONAS

Given the ability to translate between a persona-like description and items on an empirical survey, it is possible to design a survey to examine the prevalence of attributes in a written persona, or to use survey or other data to establish attributes to construct a persona. The prevalence of the persona-like description would be defined as the proportion of people in the survey whose answers are jointly consistent with the attributes as stated by the persona.

Mapping to empirical descriptions is useful for evaluating prevalence but is not required for every aspect of a persona. Attributes as given in a typical written persona do not have to be written in terms of ranges or similar characteristics of survey items; they merely map to them. It is also not necessary to map every attribute to a survey item. For example, detailed descriptions of job tasks might appear in a persona but not on a corresponding survey. Because our partial mapping establishes the joint prevalence of a subset of persona attributes, it provides an *upper estimate* of prevalence.

We investigate a general question: given the proposed formal definition of persona attributes, what is the expected prevalence of persona-like descriptions? This question can be explored analytically in terms of probability theory and investigated empirically by determining prevalence of such descriptions in actual datasets.

EXPECTED PREVALENCE RATES

To determine the analytically expected prevalence of persona-like descriptions, it is necessary to make assumptions about the prevalence of attributes that make up such descriptions. A persona attribute presents one value from a list of *n* potential values. Under the assumption of uniform distribution, the likelihood of a single draw of a given value is estimated as Pr = 1/*n*. Other distributional assumptions, such as normal, logarithmic, or Poisson fit, are possible and appropriate for some kinds of data, but because each requires some assumption or knowledge about the underlying data that is more specific than a uniform distribution (such as the order or metric qualities of the data), it is generally most consistent with the assumption of nominal data to assume uniform distribution. Likewise, for purposes of analysis, we assume that each individual persona attribute is independent of others, i.e., that the joint probability of each pair is equal to the product of their individual probabilities.

If it is assumed that each attribute is independent, then the joint probability of K attributes with uniform distribution, combined into a single persona-like description, is:

$$\Pr([persona]) = \frac{1}{n_1} * \frac{1}{n_2} * \ldots \frac{1}{n_K} = \prod_{i=1}^{K}\left(\frac{1}{n_i}\right) \quad (1)$$

Equation 1 is, however, inadequate for a general analytical solution because it relies upon knowledge of the actual number of response levels for each item. A form more appropriate for general estimation results from two additional assumptions: (1) attributes are drawn from lists with continuous ordinality in number of response levels, defined by an upper bound n_{max} and lower bound n_{min}; (2) K items are uniformly distributed in response list length. Under those assumptions, it is possible use a combinatorial formula (Rosen, 1999, chapters 2 & 7) to simplify the compound probability estimate:

$$\Pr([persona]) = \left(\frac{1}{\left(\frac{n_{max}\,!}{(n_{min}-1)!}\right)^{\frac{1}{(n_{max}-n_{min}+1)}}}\right)^{K} \quad (2)$$

For example, combining 5 items from lists of 2-4 levels estimates prevalence $\left(\frac{1}{(4*3*2)^{\frac{1}{3}}}\right)^{5} = 0.005$.

This establishes an initial *analytic* estimate of the prevalence of a persona-like description given three parameters, n_{max}, n_{min}, and K.

OBSERVED PREVALENCE RATES

When multiple attributes are combined into a single persona, how well does that description map to real people of interest? It is possible to evaluate the observed joint occurrence of persona-like descriptions (combined sets of attributes) by examining empirical datasets and determining the joint prevalence of responses.

Method

As an overview, we determined prevalence by selecting items from real survey datasets with a discrete number of observed responses (thus serving as nominal list items, as in our formal model) and finding the observed joint probability of those combined items among the real respondents. By performing this many times, across varying numbers of attributes and varying lengths of response lists, we determined the distribution of persona-like descriptions in given datasets.

The specific procedure is the following. For a given dataset, choose two properties, the number of attributes to sample (hereafter, number of *attributes*, or K) and the maximum number of nominal responses per item (hereafter, maximum number of *levels* per attribute, or n_{max}). The general algorithm is this: from the dataset, select K attributes (i.e., variables or columns) such that each has n_{max} or fewer levels. For each attribute, select a specific level (i.e., specific response). In order not to bias towards very low probabilities, sample each level with probability according to its observed prevalence. Determine how many cases (i.e., rows) match the joint set of specified responses. Iterate systematically across a range of values for K (ranging from 2-11 attributes) and n_{max} (ranging 2-11 possible responses; n_{min} =2). Sample each repeatedly to generate many persona-like descriptions and observed prevalence rates. We sampled 100 descriptions for each combination of K and n_{max}.

We applied this procedure to six survey datasets whose general properties are shown in Table 1. Each of the datasets was collected in the course of market research to identify customer segments and characteristics and thus reflects the kinds of items that might be used (in part) to describe people for personas. Five consumer surveys sampled PC users in the US (four surveys) and France (one), while a sixth survey sampled information technology administrators in the US. Space does not allow full description of survey items, but each included 70 or more nominal variables (see Table 1). Complete variable lists are available from the first author.

Two simulated datasets were generated using multivariate data simulation procedures (Revelle, 2005), with varying characteristics. Each had 50 variables loading on 10 factors, and was reduced from simulated "observed" scores with error terms to nominal variables uniformly cut into 2 to 11 nominal categories. For "Simulated 1," factor loadings were set to vary from 0.5 to 0.9 for each variable on its primary loading factor, and -0.2 to 0.2 on other factors, while factors were set to have low covariance (ranging -0.01 to 0.01) and high unique variance (0.8). "Simulated 2" had stronger primary factor loadings (ranging 0.6 to 0.9) but also higher secondary factor loadings (0.0 to 0.1) and independent factors (1.0 unique variance on the first 9 factors; 0.5 on the 10th, with 0.0 covariance between factors). These constructed datasets are representative of reliable, strongly factorial data, such as carefully constructed intelligence or personality measures.

For each dataset, all 100 possible combinations of K attributes & n levels were each sampled 100 times, yielding a total of 10000 randomly sampled persona-like descriptions. For each trial (i.e., each randomly generated persona-like description), we recorded the observed prevalence within the dataset. Additionally, for each of the 10000 runs in each dataset, we computed the analytically expected prevalence as given by Equation 2, based on the value of K and n_{max} for that trial (n_{min} was always set to 2 levels).

All statistical and sampling models were developed and run using the R statistics environment, version 2.6.1 (R Development Core Team, 2007).

Results

The *performance* of a given description is the prevalence rate of people that it matches in a dataset. Across 10000 runs, performance can be assessed at different likelihood points, such as the median performance, best single performance, and so forth. We selected the 99th percentile in order to examine the performance of near-optimal descriptions that are not extreme outliers. Figure 1 plots the 99th percentile prevalence of generated persona-like descriptions for each dataset, for each number of attributes, 2 .. K. The 99th percentile is the point at which 1% of the generated descriptions performed equivalently or better, and 99% performed worse than the shown prevalence rate. Figure 2 provides a close-up view of performance rates at the 99th percentile when 7 to 11 attributes are considered.

In all cases, the observed prevalence of persona-like descriptions declines rapidly as the number of attributes increases. In Figure 2, the 99th percentile level fails to match anyone when 9 or more attributes are combined in 5 out of 6 survey datasets – including all consumer datasets. Only in the IT dataset did the 99th percentile performance level match more than 0.3% of real respondents with 7 or more attributes.

Accuracy of the analytic prediction of declining prevalence can be evaluated by examining the correlation between observed values and the model's predicted values. Note that prevalence rates are not expected to be normally distributed; the predicted values decrease geometrically with K. Therefore, interpretation of correlation strength from raw prevalence scores alone would be misleading, as the values of r cannot be interpreted according to rules of thumb that apply to normal distributions (Cohen, 1988; Box & Cox, 1964).

Table 1: Properties of datasets

Name	Source	Cases (N)	Variables (#)	Variables (#), n_{max}=2 levels	Variables (#), 2≤n_{max}≤11 levels
IT survey	Authors, 2001	1915	969	316	725
Consumer US 1	Unrelated research team, 2005	10307	729	36	682
Consumer US 2	Authors, 2006	916	320	77	263
Consumer US 3	Authors, 2007	786	169	14	70
Consumer US 4	Authors, 2007	665	185	79	159
Consumer France	Authors, 2007	268	114	26	112
Simulated 1	Simulated	2000	50	6	50
Simulated 2	Simulated	2000	50	9	50

Figure 1. Prevalence rates by dataset and number of attributes at 99th percentile of generated descriptions

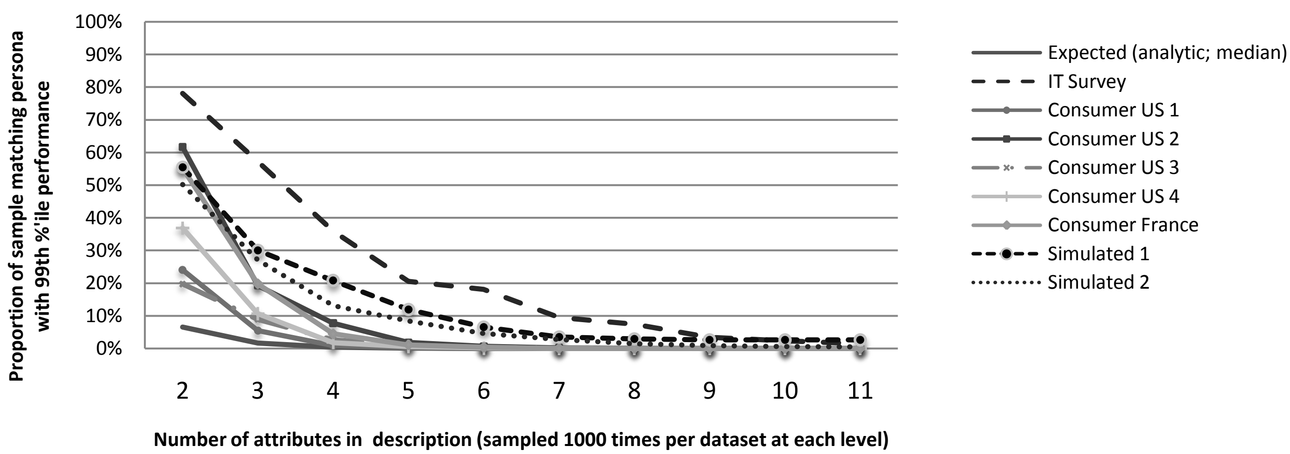
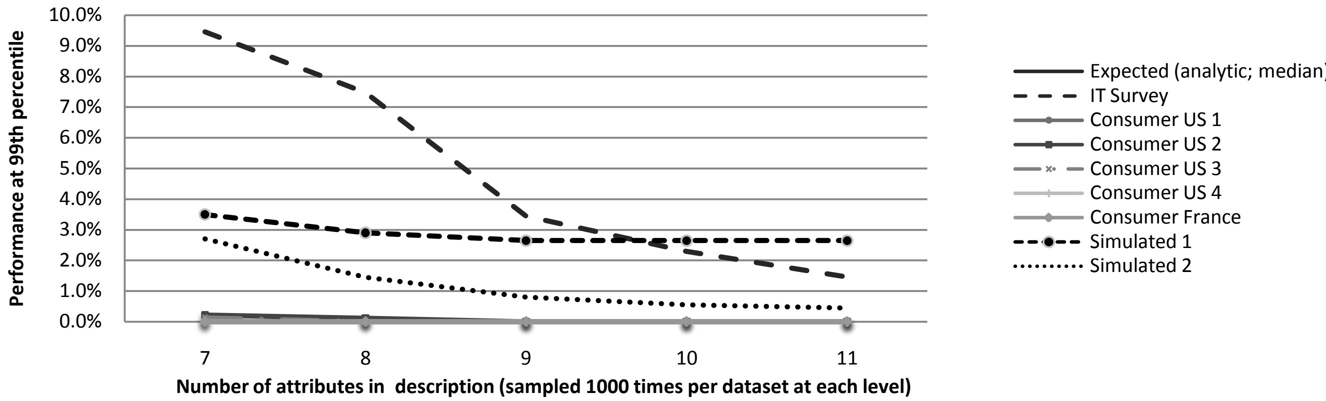


Figure 2. Close-up on prevalence by dataset and number of attributes at 99th percentile performance, for 7-11 attributes



Correlation between predicted and observed prevalence rates is shown in Table 2, listing correlation coefficients of both raw values, and, more interpretably, values optimally transformed to multivariate normality by the Box-Cox procedure (Box & Cox, 1964; Fox, 2006; offset of +0.01 to handle zero values). Pearson's $r(9998)$ values range from 0.394 to 0.713 for multinormal results from survey datasets, and 0.857 to 0.869 for simulated data (all $p < 0.001$), showing strong to very strong correlation between the predicted and observed values in all datasets (Cohen, 1988).

Table 2. Correlation of observed vs. predicted prevalence

Dataset (N=10000 trials each)	***r*, raw observed vs. predicted prevalence**[a]	***r*, transformed observed vs. predicted prevalence**[a]
IT survey	0.636	0.713
Consumer US 1	0.176	0.394
Consumer US 2	0.201	0.493
Consumer US 3	0.242	0.540
Consumer US 4	0.404	0.484
Consumer France	0.160	0.471
Simulated 1	0.728	0.857
Simulated 2	0.759	0.869

[a] all $p < 0.001$, $df = 9998$ (exact $p < 10^{-15}$ in all cases)

DISCUSSION

We developed a formal definition of at least some of the attributes expressed by personas. Using that formal model, the predicted prevalence rates of persona-like descriptions showed strong correlation to observed prevalence in six real and two simulated datasets. As predicted, the observed prevalence rate of persona-like descriptions declined rapidly as attributes were added.

The main implication is that a persona-like description with more than a few attributes cannot be assumed to describe many actual people. Thus, if a persona's authors choose to claim that a persona conveys empirical information about people, they should also establish empirical evidence for that claim.

One way to support a persona's empirical claims would be to report survey items that map to its attributes along with data on their empirically observed multivariate prevalence rates. It is important to note that examining prevalence rates of individual attributes would be insufficient. Rather, the persona's authors should determine the multivariate incidence of attributes in combination.

The present sampling and generation method has a potentially useful application: although it does not represent a method for constructing personas, it could be used to determine whether a persona is performing better than chance. This method could be used to find the prevalence of an authored persona and determine whether it accounts for more people than randomly generated descriptions at some criterion level (e.g., 50th, 90th, or 99th percentiles).

The present work does not address the *utility* of personas. The personas method is claimed to lead to positive results (e.g., Pruitt & Adlin, 2006, p. 3) and it is possible that the method could be useful for inspirational purposes even if the information claims are wrong, i.e., even if personas do not actually describe people. Assessing the question of utility would require a different research model, and it might be difficult to implement as a well-controlled study.

Finally, attributes and values could be sampled under assumptions different than the ones we selected. By employing an empirical sampling method akin to bootstrap estimation using real datasets, our results should be robust. A generalization of the procedure could be used precisely to estimate distributional parameters in a multivariate dataset (Efron & Tibshirani, 1994).

CONCLUSION

As shown, generated descriptions with many attributes are unlikely to identify many people. A final question is whether a *designed* persona would identify many more people. It may be possible to develop a persona with higher prevalence than our model would estimate. An interesting research agenda would be to determine whether such performance could be achieved systematically. In order to exceed the performance rates noted here, such a constructed persona would have to surpass the observed performance of the top 1% of generated descriptions.

Our key assertion is that persona authors should expect a description with many attributes to have very low prevalence, and therefore the informational content and population relevance of personas must be *assessed* and not simply assumed. Personas need empirical evidence to substantiate claims that they present factual information about groups of people. We hope that the HCI community will undertake further systematic exploration of this widely used technique. Until that time, theoretical methods, such as the model which has been presented here, may provide the best estimates of the ability of personas to describe real world sets of people.

REFERENCES

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B,* 26, 211–246.

Brechin, E. (2002). Reconciling market segments and personas. Cooper Design. Retrieved November 28, 2007 from http://www.cooper.com/newsletters/2002_02/reconciling_market_segments_and_personas.htm

Chapman, C. N., & Milham, R. P. (2006, October). The personas' new clothes: methodological and practical arguments against a popular method. *Proceedings of the Human Factors and Ergonomics Society (HFES) 50th Annual Conference*, San Francisco, CA.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, A. (1999). *The inmates are running the asylum*. New York: Macmillan.

Cooper, A., Reimann, R., & Cronin, D. (2007). *About face 3: The essentials of interaction design*. Indianapolis: Wiley.

Courage, C., & Baxter, K. (2005). *Understanding your users: A practical guide to user requirements*. San Francisco: Morgan Kaufmann.

Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Fox, J. (2006) car: Companion to applied regression [Computer software]. R package, version 1.2-1, http://socserv.socsci.mcmaster.ca/jfox/

Holtzblatt, K., Wendell, J. B., & Wood, S. (2005). *Rapid contextual design: A how-to guide to key techniques for user-centered design*. San Francisco: Morgan Kaufmann.

Mulder, S., & Yaar, Z. (2006). *The user is always right: A practical guide to creating and using personas for the web*. Berkeley, CA: New Riders.

Pruitt, J. & Adlin, T. (2006) *The persona lifecycle: keeping people in mind throughout product design*. San Francisco: Morgan Kaufmann.

Pruitt, J. & Grudin, J. (2003). Personas: Practice and Theory. Proceedings from DUX 2003: *Designing for User Experience Conference 2003*.

R Development Core Team (2007). R: A language and environment for statistical computing [Computer software]. Version 2.6.1. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Revelle, W. (2005). Multivariate data construction routines [Computer software]. Retrieved December 21, 2007 from http://personality-project.org/r/r.datageneration.html.

Rosen, K.H., ed. (1999) *Handbook of discrete and combinatorial mathematics*. Boca Raton: CRC.