Background and Methodology

The National Cancer Institute's (NCI) Research-Based Web Design and Usability Guidelines project began in March of 2000. Since that time, each guideline presented in this book has undergone an extensive internal and external review. The process used to create the *Guidelines* is presented here.

Step 1: Creating the Initial Set of Guidelines

The NCI's Communication Technologies Branch (CTB) needed to develop a set of guidelines that would help designers build websites that are based on the best available research. The initial set of guidelines were drawn from existing Web design guideline and style guides, published research articles, research summaries, publicly available usability test reports, and lessons learned from in-house usability tests. This effort resulted in more than 500 guidelines.

Step 2: Reviewing the Initial Set of Guidelines

The initial set of 500 guidelines was far too many for website designers to use effectively. CTB initiated an internal review process to:
- Identify and combine similar guidelines;
- Identify and resolve guidelines that conflicted with each other; and
- Reword unclear guidelines.

This internal review was conducted by CTB staff and consultants. Each of the reviewers had experience in website design, usability engineering, technical communication, software design, computer programming and/or human-computer interaction. The internal review reduced the initial set of guidelines to 398.

Step 3: Determining the "Relative Importance" of Each Guideline

To determine the 'Relative Importance' of each guideline, sixteen external reviewers were recruited. Half of these reviewers were website designers and half were usability specialists. Reviewers evaluated each guideline and then answered the question, "How important is this guideline to the overall success of a website?" by assigning a score from a scale that ranged from "Extremely Important" to "Not Important."

Step 4: Validating the Initial "Relative Importance" Ratings

After the initial review by the 16 website practitioners (designers and usability specialists), the set of guidelines was reduced to 287. Those guidelines that were rated as having little importance to the success of a website were eliminated. Many guidelines were edited and clarified based on feedback from the reviewers. Also, a few new guidelines were added as new research was gathered.

To validate the "Relative Importance" ratings, the same 16 evaluators were asked to confirm or modify their previous ratings with knowledge of their own and the average rating from the previous review.

Step 5: Determining the "Strength of Evidence" for Each Guideline

The next step was to generate a reliable 'Strength of Evidence' rating for each guideline. To do this, CTB recruited a group of eight researchers from a variety of fields—including usability, user experience, documentation, computer science, and cognitive psychology—that have an influence on Web design. These reviewers were all published researchers with doctoral degrees, experienced peer reviewers, and knowledgeable of experimental design.

Developing the "Strength of Evidence" ratings for each guideline was conducted in three parts. In Part One, reviewers were asked to classify each guideline as having "strong," "weak," or "no" research evidence to support it. The goal was to determine which guidelines had no research evidence so that they could be pulled out, and hence, help reviewers focus on rating the remaining set. Reviewers also were asked to provide new sources of evidence for each guideline (if available).

Based on the results of Part One, the project team learned that there was very little agreement on what constitutes "strong," "weak," or "no" research evidence.

Therefore, the project team planned Part Two to generate a common framework among the reviewers.

Part Two had the reviewers attend a one-day meeting and agree on the following scale for rating the "Strength of Evidence" for each guideline.

5 – Strong Research Support ①②③④⑤
- Cumulative and compelling, supporting research-based evidence
- At least one formal, rigorous study with contextual validity
- No known conflicting research-based findings
- Expert opinion agrees with the research

4 – Moderate Research Support ①②③④◯
- Cumulative research-based evidence
- There may or may not be conflicting research-based findings
- Expert opinion
  - Tends to agree with the research, and
  - A consensus seems to be building

3 – Weak Research Support ①②③◯◯
- Limited research-based evidence
- Conflicting research-based findings may exist
  - and/or -
- There is mixed agreement of expert opinions

2 – Strong Expert Opinion Support ❶❷〇〇〇
- No research-based evidence
- Experts tend to agree, although there may not be a consensus
- Multiple supporting expert opinions in textbooks, style guides, etc.)
- Generally accepted as a 'best practice' or reflects 'state of practice'

1 – Weak Expert Opinion Support ❶〇〇〇〇
- No research-based evidence
- Limited or conflicting expert opinion

The reviewers also agreed upon a set of categories to classify the many sources that had been collected. The reviewers assigned each reference to one of the following categories:

- Rigorous observational study (e.g., ethnographic evaluation)
- Hypothesis-oriented experiment
- Model-based evaluation
- Expert opinion with no or few references
- Reference-base literature review, chapter in a book, or meta-analysis
- Survey
- Textbook with many references
- Usability test results or summary of several usability tests (e.g., lessons learned)
- Exploratory study (e.g., "How long will people wait for a page to download?")

Part Three had reviewers evaluate the available evidence for each guideline, and then assign a rating based on the 5-point scale described above. Because of the activities in Part Two, agreement among reviewers in classifying the evidence for each guideline substantially increased.

Step 6: Finding Graphic Examples for the Guidelines
To ensure that users clearly understand the meaning of the guideline, the project team identified and reviewed several possible examples for each guideline, and selected the strongest examples.

Step 7: Grouping, Organizing, and Usability Testing the Guidelines
To ensure that the information about specific Web design issues is easy to find, a group of twenty website designers were asked to participate in a formal "grouping" of the guidelines by using a card-sorting exercise. Each of the twenty individuals put the guidelines into groups that reflected how they think about Web design issues, and then provided a name for each group. Data from this exercise was analyzed with specially developed software and formed the chapters of this book.

Several draft page layouts in print format were developed for this book. These drafts were usability tested to determine how best to facilitate readers' ability to locate and understand information on a page. These findings, as well as readers' preferences, served as the basis for the final page layout.