National Cancer Institute

Utilizing Informatics to Foster Collaboration and Networking; Challenges and Opportunities

Synergizing Epidemiologic Research on Rare Cancers

May 10-11, 2007

Bethesda, MD

National Cancer Institute

The Current World

- Islands of biomedical information
- No automated infrastructure to visit / share/ collaborate/ discover

National Cancer Institute

Web 1.0

- The **World Wide Web** (or the "**Web**") is a system of interlinked, hypertext documents that runs over the Internet. With a Web browser, a user views Web pages that may contain text, images, and other multimedia and navigates between them using hyperlinks. The Web was created around 1990 by the Englishman Tim Berners-Lee and the Belgian Robert Cailliau working at CERN in Geneva, Switzerland. Since then, Berners-Lee has played an active role in guiding the development of Web standards (such as the markup languages in which Web pages are composed), and in recent years has advocated his vision of a Semantic Web.

National Cancer Institute

Web 1.0 to 2.0

- The transition of Web sites from isolated information silos to sources of content and functionality, thus becoming computing platforms serving web applications to end-users
- A social phenomenon embracing an approach to generating and distributing Web content itself, characterized by open communication, decentralization of authority, freedom to share and re-use, and "the market as a conversation"

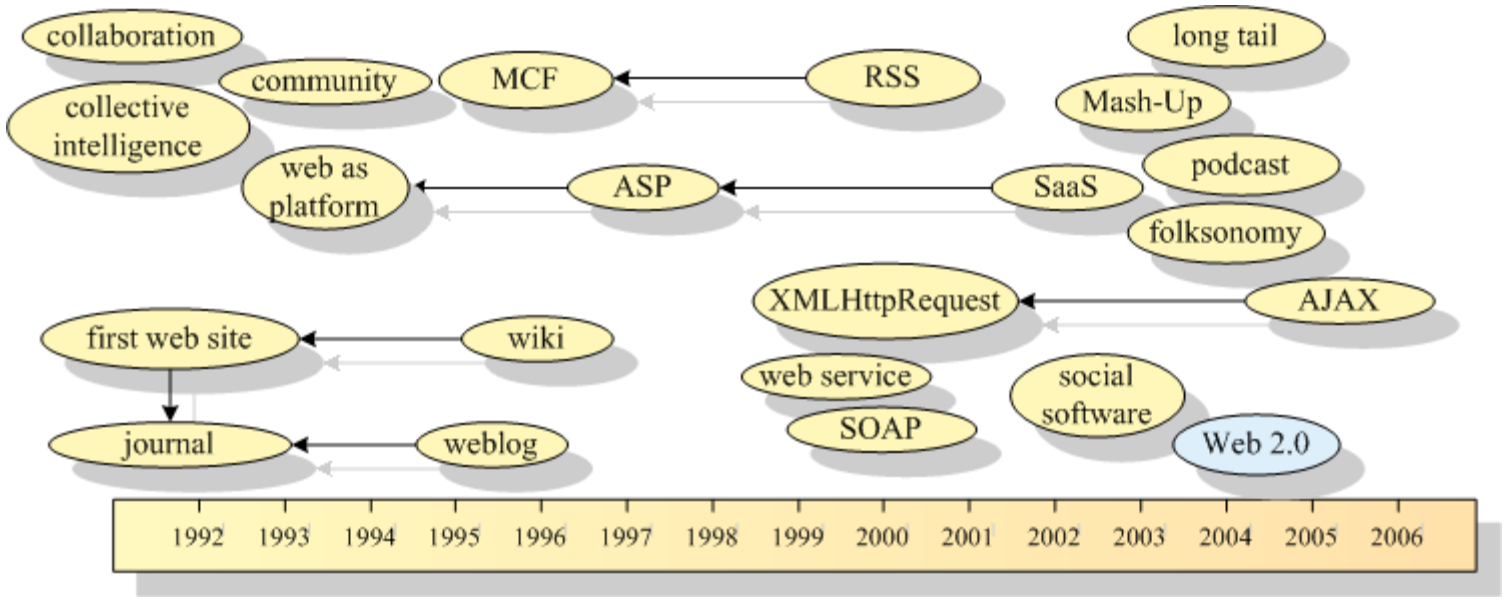
National Cancer Institute

Web 2.0

- **Web 2.0**, a phrase coined by O'Reilly Media in 2004,[1] refers to a perceived second-generation of Web based communities and hosted services — such as social networking sites, wikis and folksonomies — that facilitate collaboration and sharing between users. O'Reilly Media titled a series of conferences around the phrase, and it has since become widely adopted.

National Cancer Institute

Web 2.0



National Cancer Institute

The Believers

- Wikinomics: How Mass Collaboration Changes Everything, by Don Tapscott (Author), Anthony D. Williams (Author)

- The emerging Web 2.0 social software: an enabling suit of sociable technologies in health and health care education.
  - *Health Information and Libraries Journal, 24, pg 2-23, 2007*

National Cancer Institute

"Danger Will Robison"

said the Robot in Lost in Space

- WekiMedia
  - Nature Medicine Vol 13 #3 March 2007
    - 22 year old is major editor adding to medical wekipedia
- Lack of credentialing
  - Who are you? What gives you the right or knowledge?
- Group think
  - The madness of crowds
    - The human genome sequence
    - dot com's
    - Investment banking
  - The world is flat

National Cancer Institute

Web 3.0

- The **semantic web** is an evolving extension of the World Wide Web in which web content can be expressed not only in natural language, but also in a form that can be understood, interpreted and used by software agents, thus permitting them to find, share and integrate information more easily.[1] It derives from W3C director Tim Berners-Lee's vision of the Web as a universal medium for data, information, and knowledge exchange.
- At its core, the semantic web comprises a philosophy,[2] a set of design principles,[3] collaborative working groups, and a variety of enabling technologies. Some elements of the semantic web are expressed as prospective future possibilities that have yet to be implemented or realized.[4]
- Transforming the Web into a database
  - The first step towards a "Web 3.0" is the emergence of "The Data Web" as structured data records are published to the Web in reusable and remotely queryable formats, such as XML, RDF and microformats.
    - NSF Cyberinfrastructure
    - NCI caBIG

National Cancer Institute

The GRID is Web 3.0

- **Grid Computing** is an emerging computing model that treats all resources as a collection of manageable entities with common interfaces to such functionality as lifetime management, discoverable properties and accessibility via open protocols.

- Like the Internet, grid computing evolved from the computational needs of "big science". The Internet was developed to meet the need for a common communication medium between large, federally funded computing centers. These communication links led to resource and information sharing between these centers and eventually to provide access to them for additional users. Ad hoc resource sharing 'procedures' among these original groups pointed the way toward standardization of the protocols needed to communicate between any administrative domain. The current grid technology can be viewed as an extension or application of this framework to create a more generic resource sharing context.
- The ideas of the grid (including those from distributed computing, object oriented programming, cluster computing, web services and others) were brought together by Ian Foster, Carl Kesselman and Steve Tuecke, widely regarded as the "fathers of the grid[1]." They led the effort to create the Globus Toolkit incorporating not just CPU management (examples: cluster management and cycle scavenging) but also storage management, security provisioning, data movement, monitoring and a toolkit for developing additional services based on the same infrastructure including agreement negotiation, notification mechanisms, trigger services and information aggregation. In short, the term grid has much further reaching implications than the general public believes. While Globus Toolkit remains the de facto standard for building grid solutions, a number of other tools have been built that answer some subset of services needed to create an enterprise grid.

National Cancer Institute

"Populomics"

New eHealth solutions may soon permit the real-time integrative utilization of vast amounts of behavioral-, biological-, and community-level information in ways <u>not previously possible</u>. Behavioral algorithms and decision support tools for scientists could facilitate the analysis and interpretation of population level data to enable the development of "community (population) arrays" or community-wide risk profiles, which in turn could form the foundation of a new "populomics."

-- Gibbons, 2005

National Cancer Institute

Interoperability

ability of a system to access and use the parts or equipment of another system

Syntactic interoperability

Semantic interoperability

National Cancer Institute

Obstacles to Population Health

- Absence of shared Ontologies and Vocabularies
  - an **ontology** is a data model that represents a set of concepts within a domain and the relationships between those concepts. It is used to reason about the objects within that domain.
- Redundancy and lack of coordination in field-initiated research
- Disincentives for "team science"
- Lack of a coordinated "presentation layer" to policy makers, planners
- Excessive time lag from discovery to delivery (AHRQ estimates 18 years)

National Cancer Institute

Standards and Vocabularies are Crucial Requirements

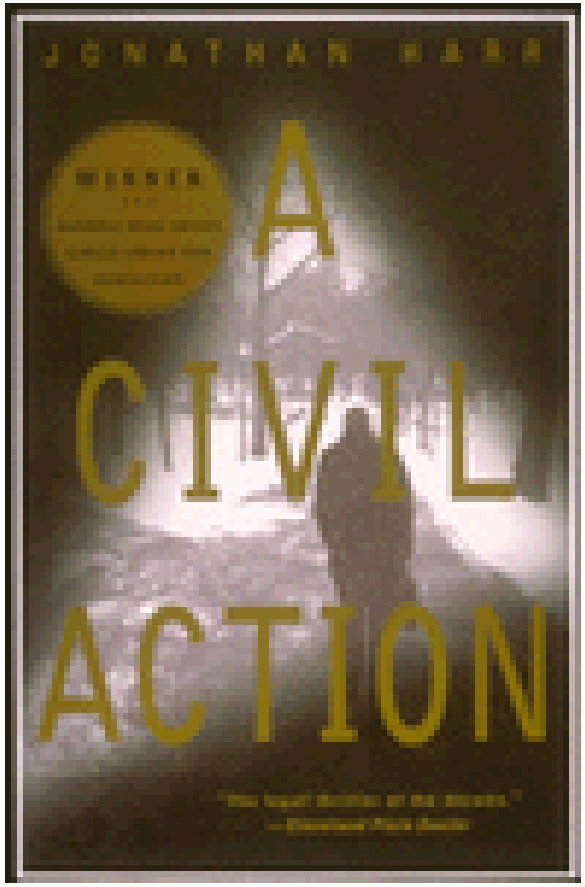
Without Standards…..

- Population data is non-comparable
- Systems cannot interchange data
- Secondary uses (Research) are not possible
- Real Time Health Surveillance is not possible
- Linkage to decisions support is not possible
- Health outcomes cannot expand to individuals and policy makers

National Cancer Institute

Possible Relative Real Time Example

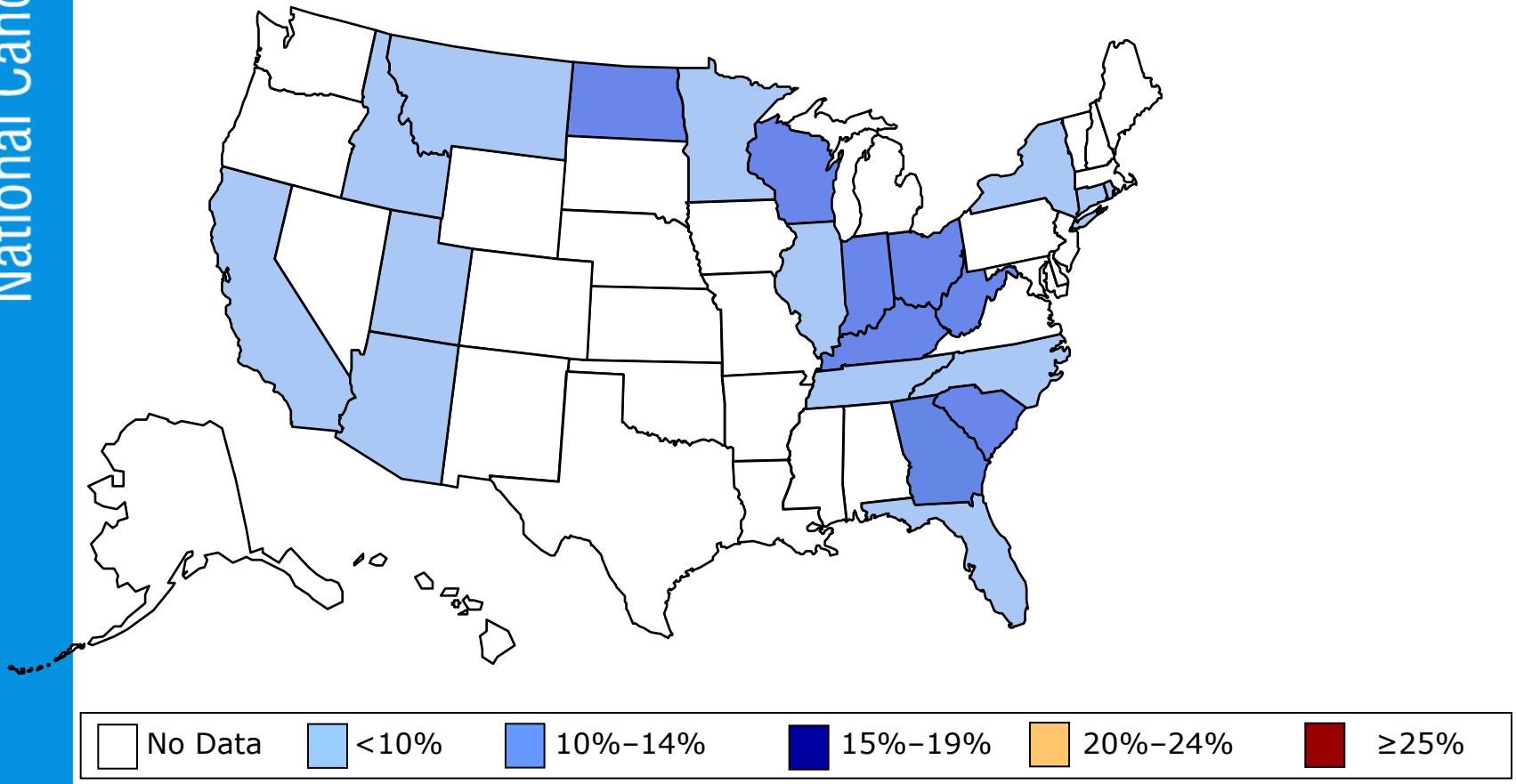
- What about Woburn MA and rare childhood leukemia's



National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1985

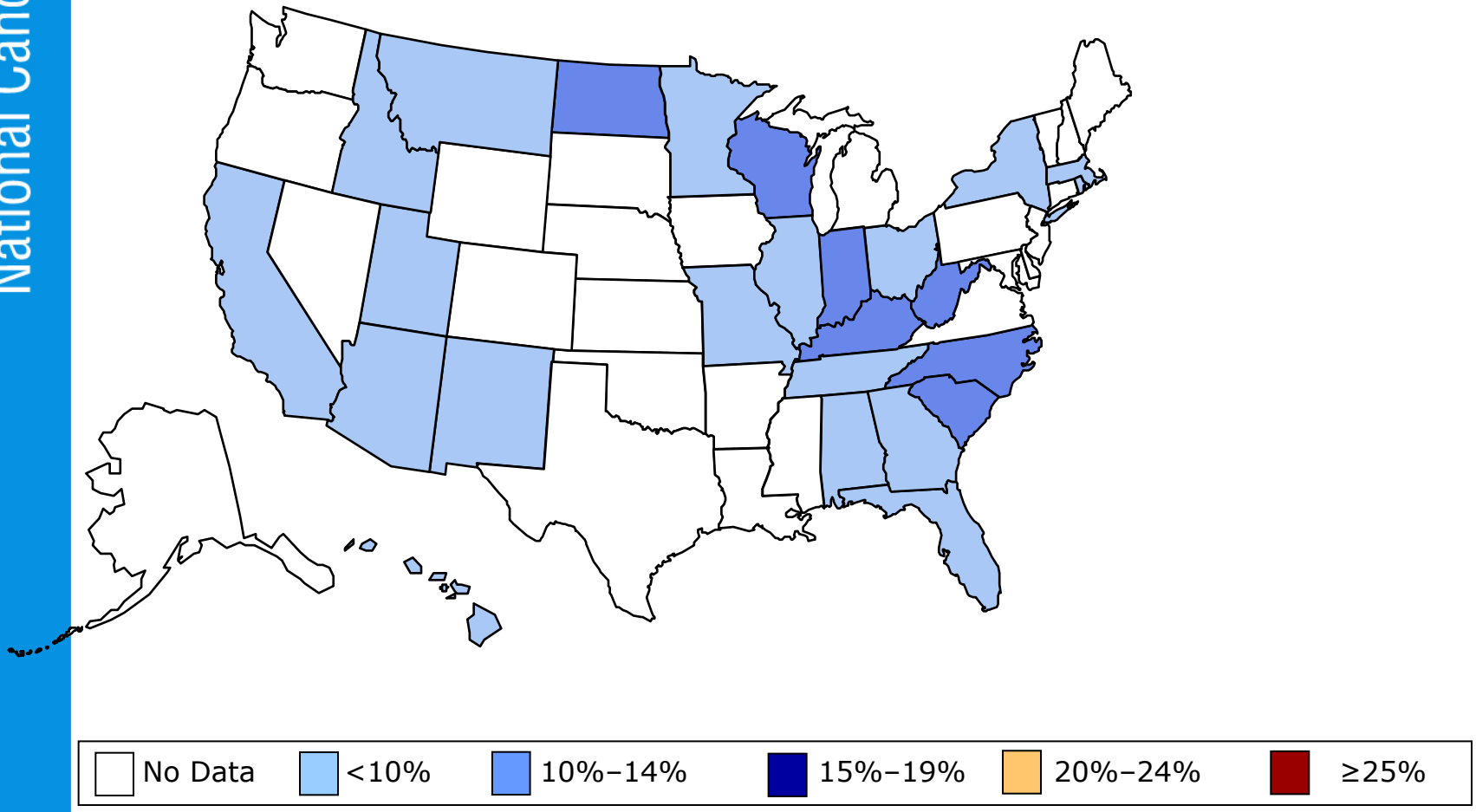


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1986

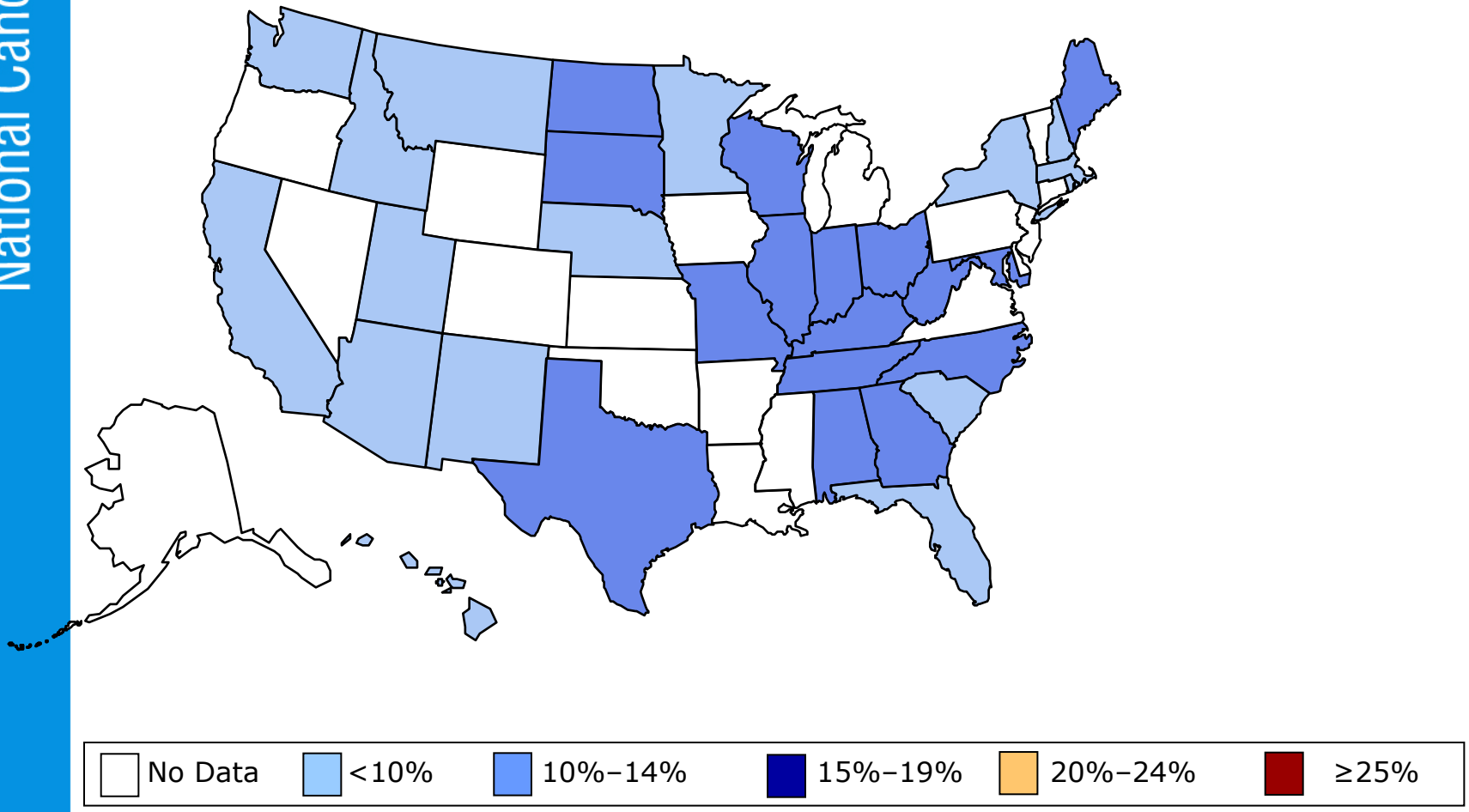


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1987

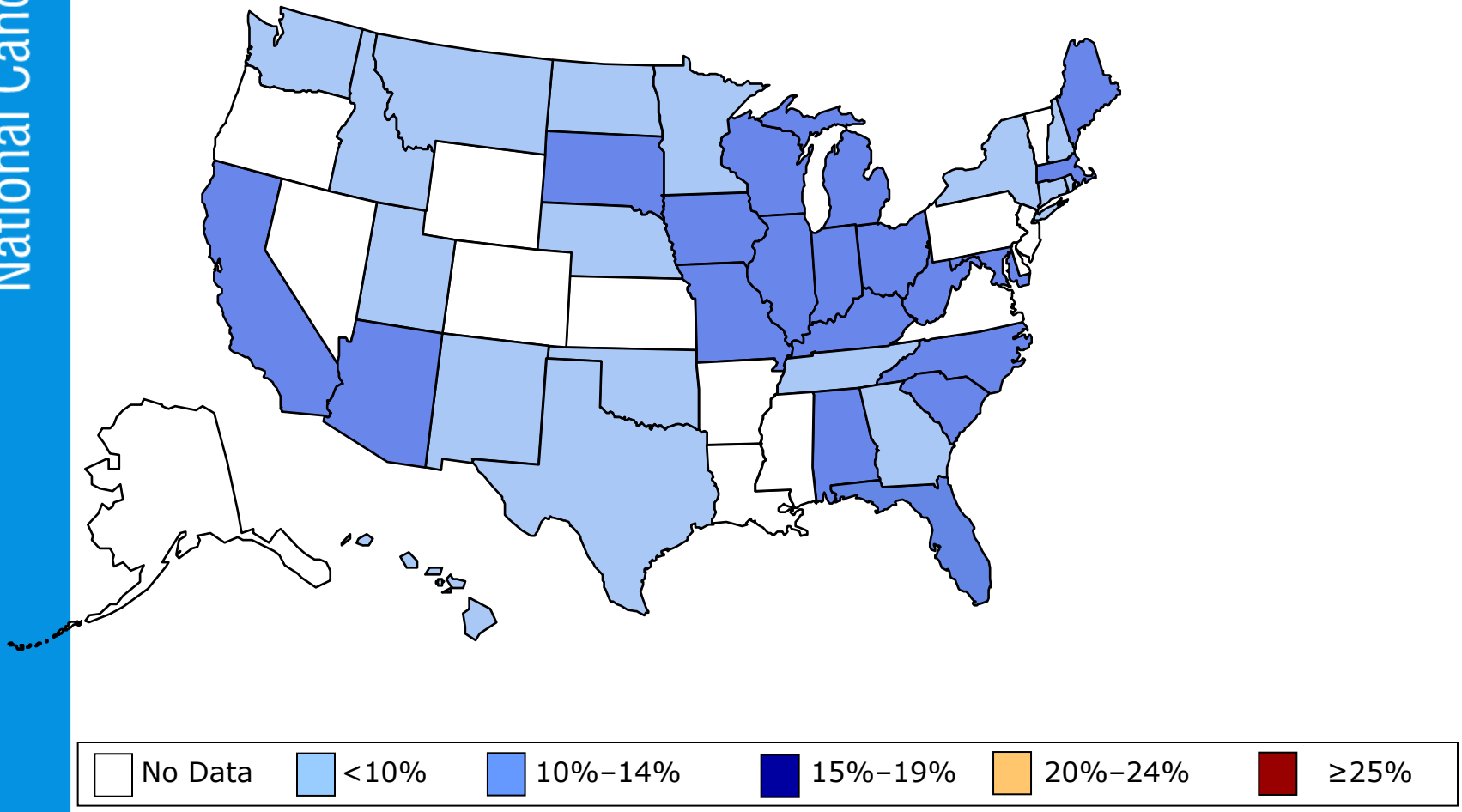


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1988

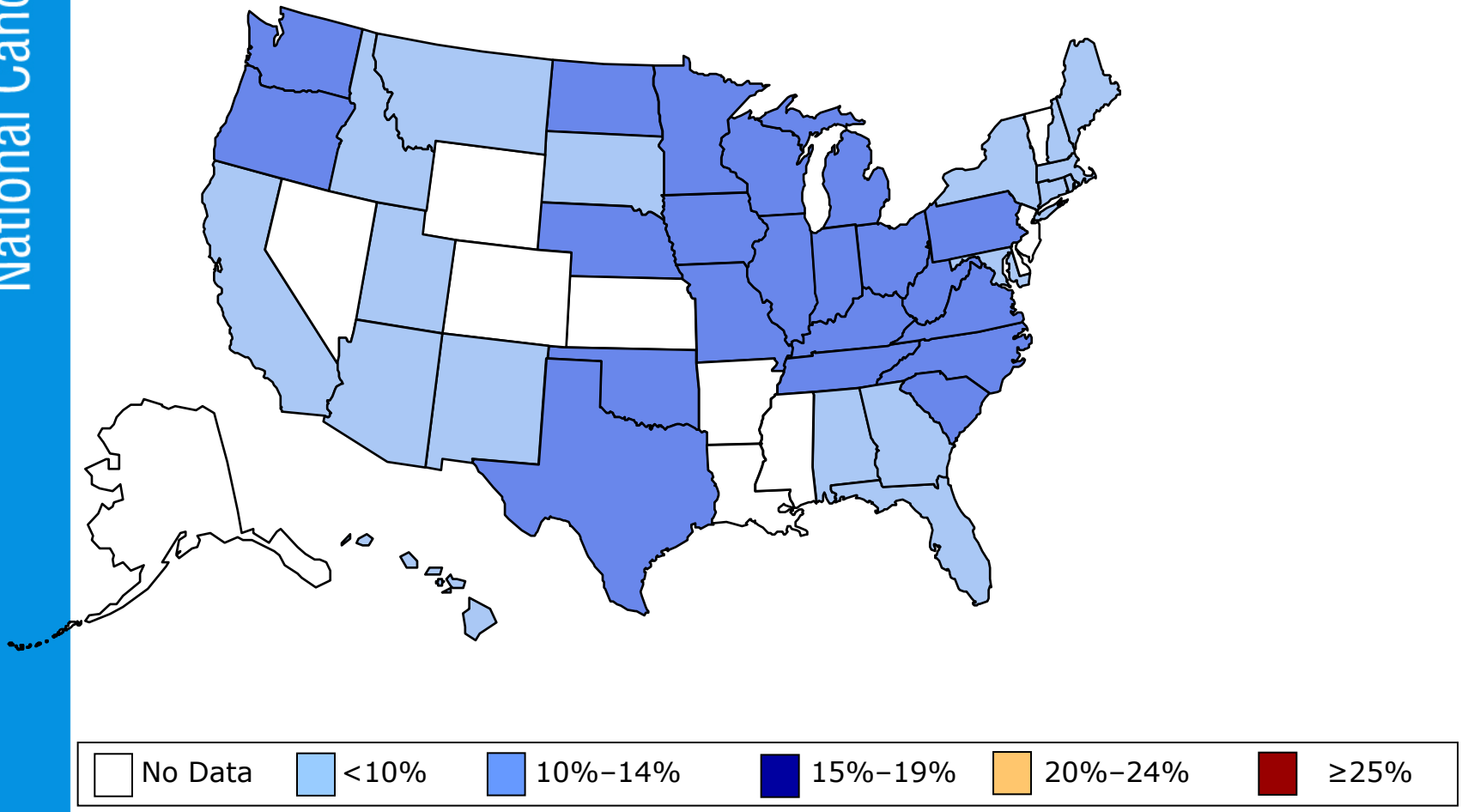


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

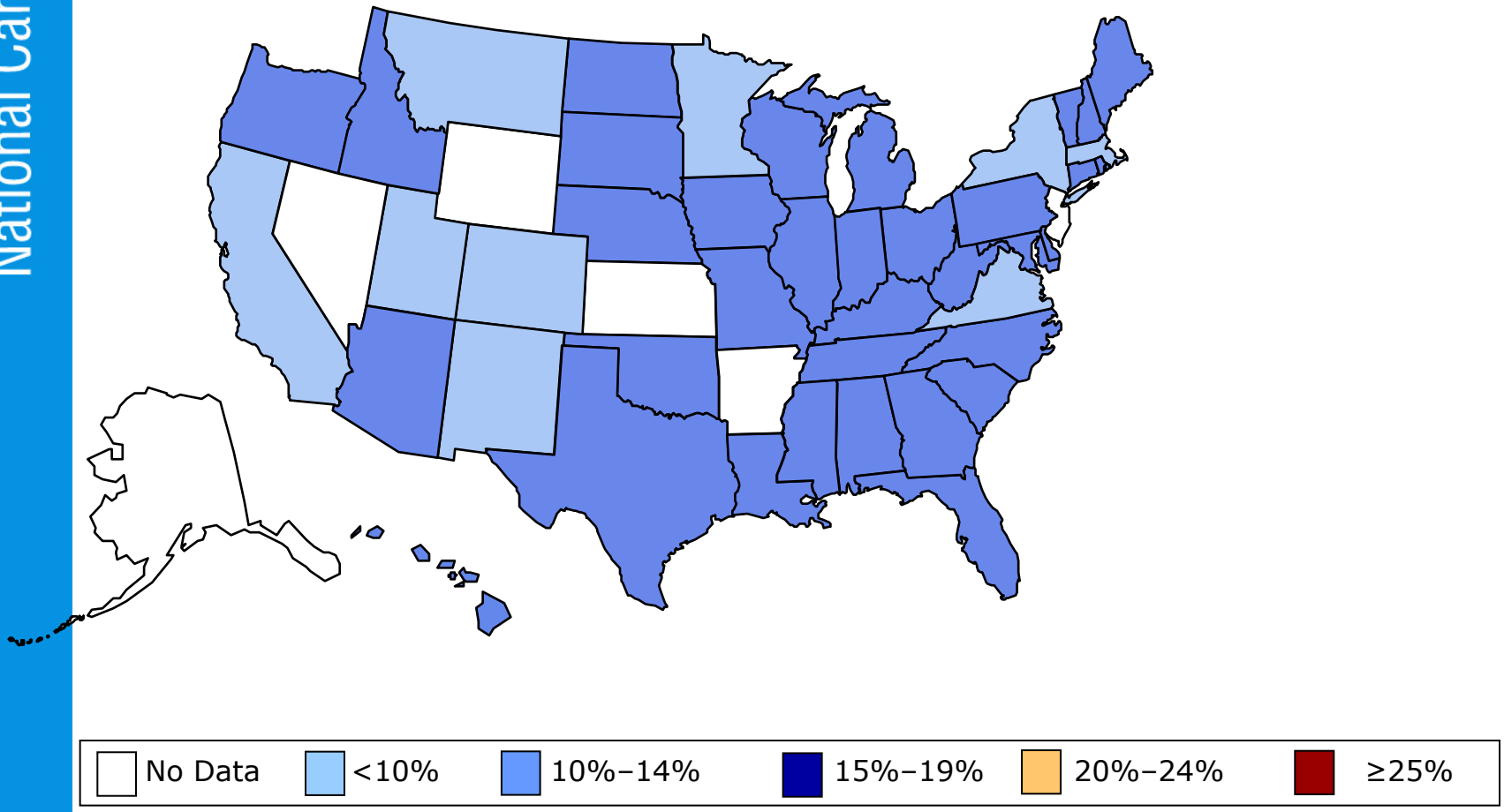
BRFSS, 1989



(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults
BRFSS, 1985

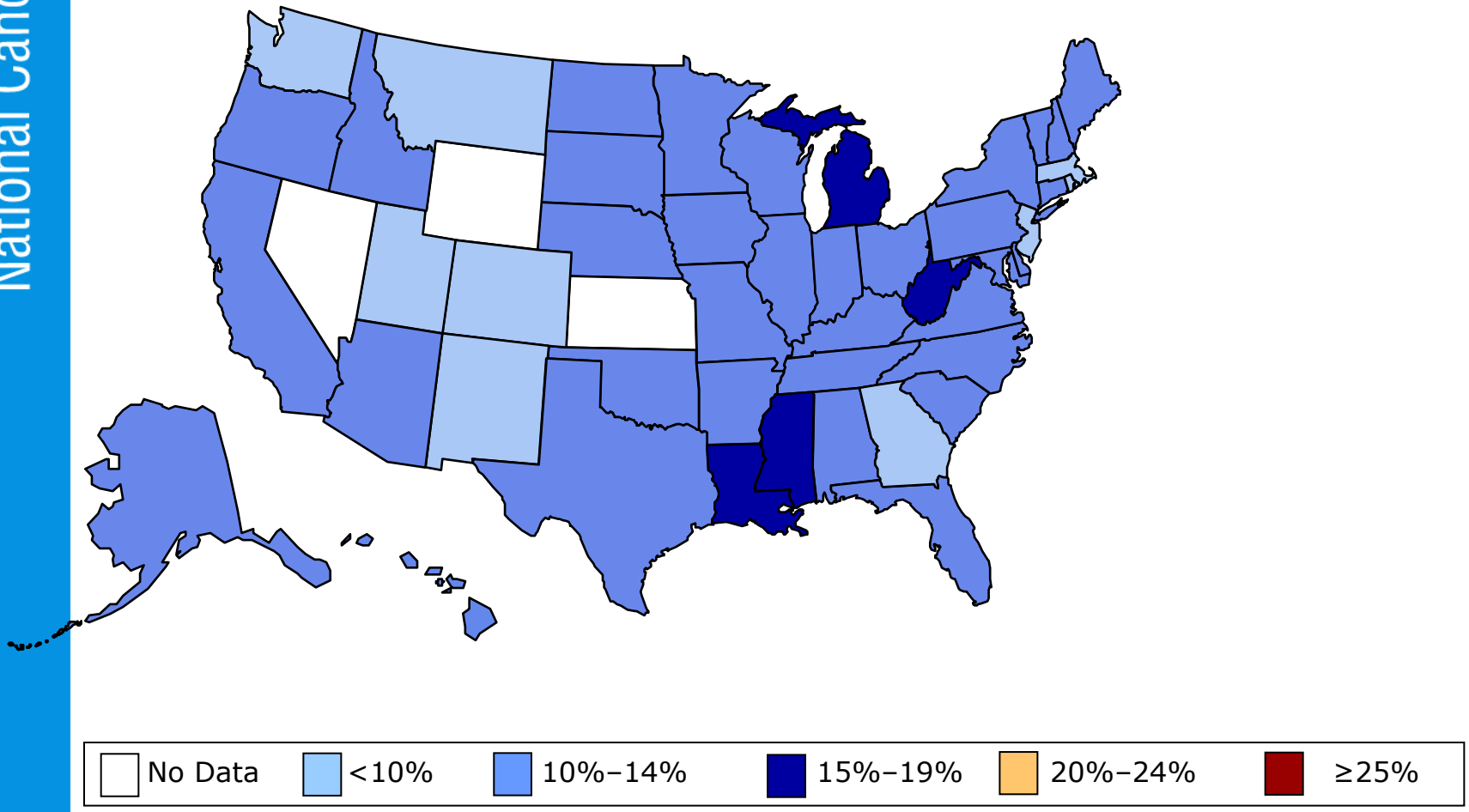


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1991

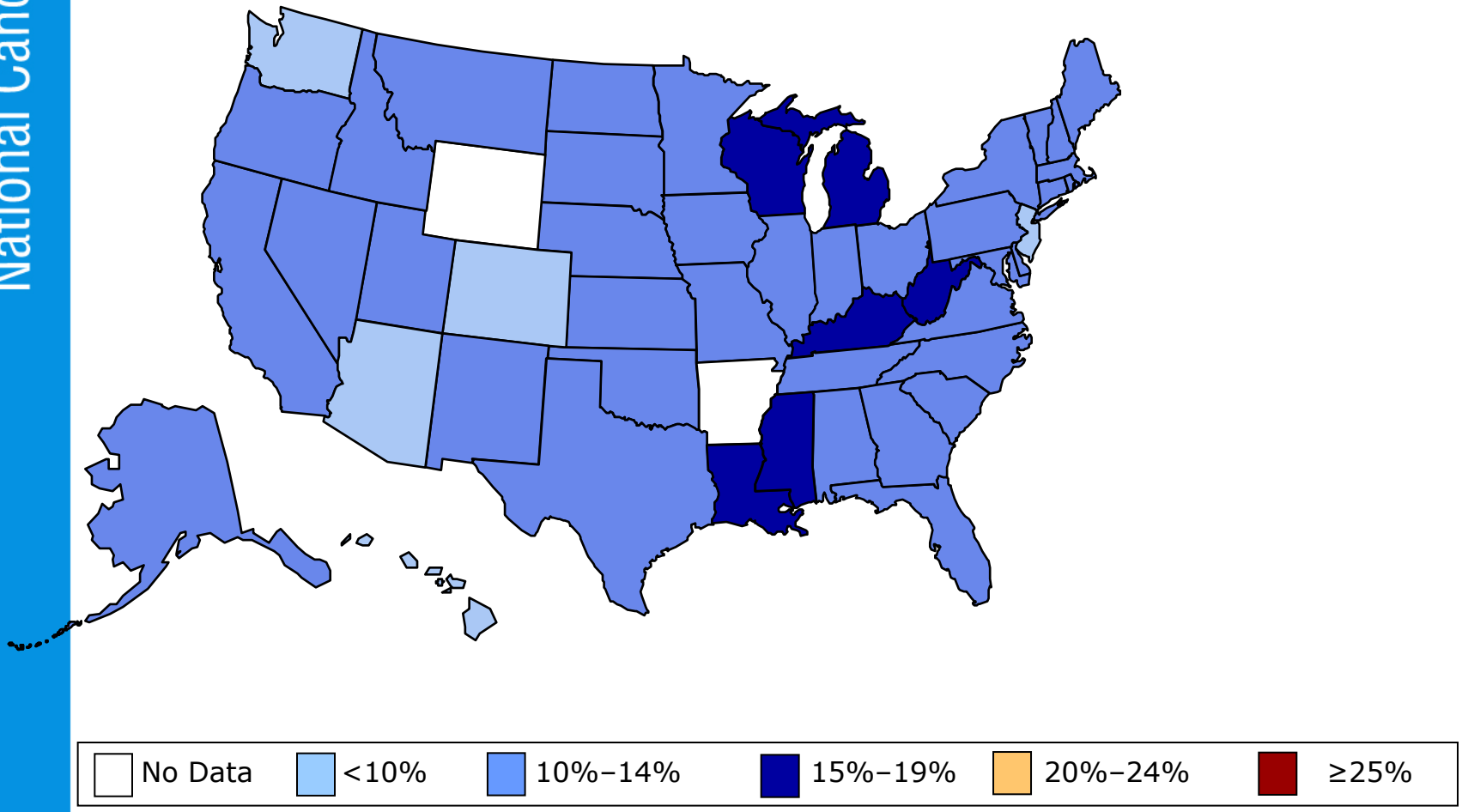


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1992

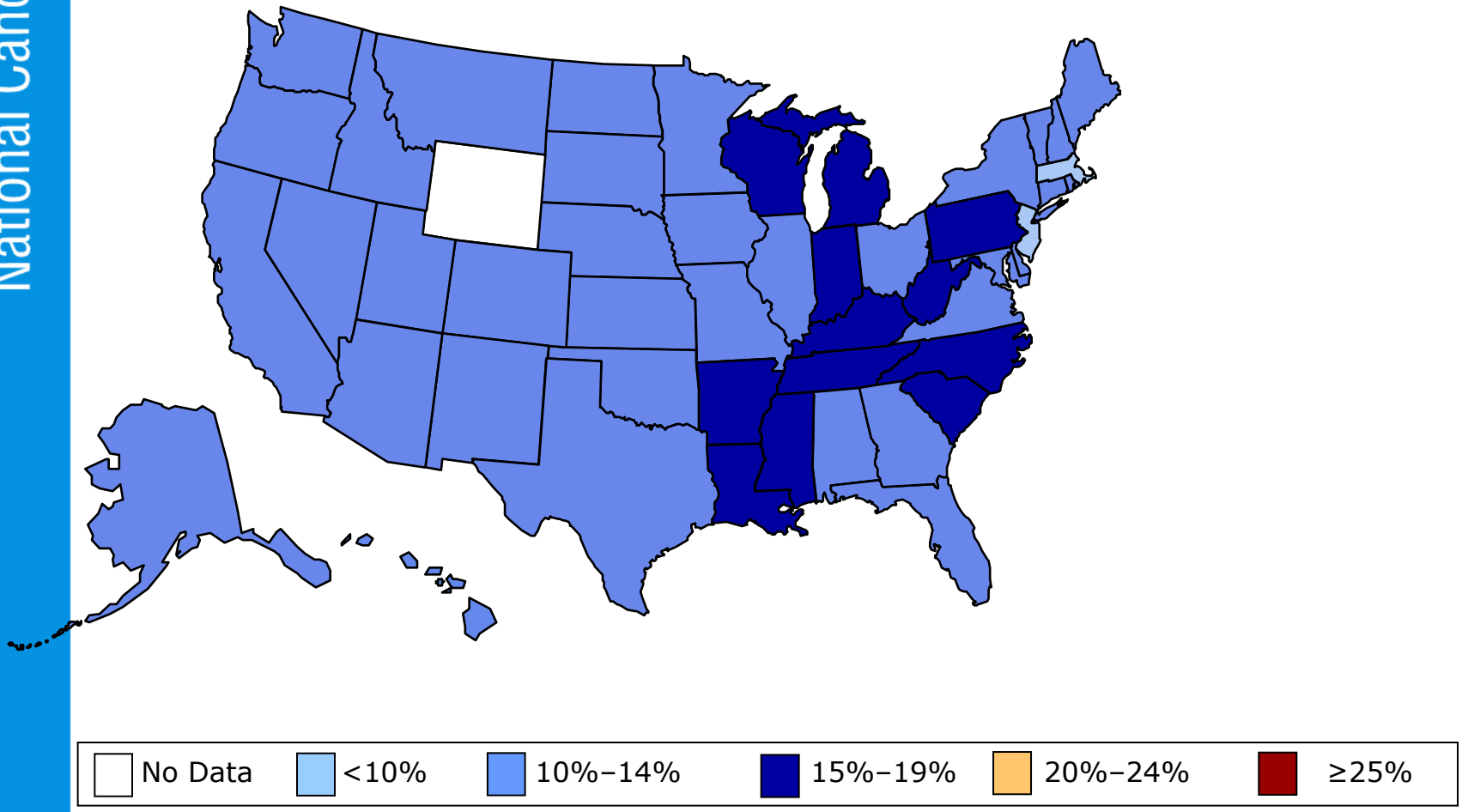


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1993

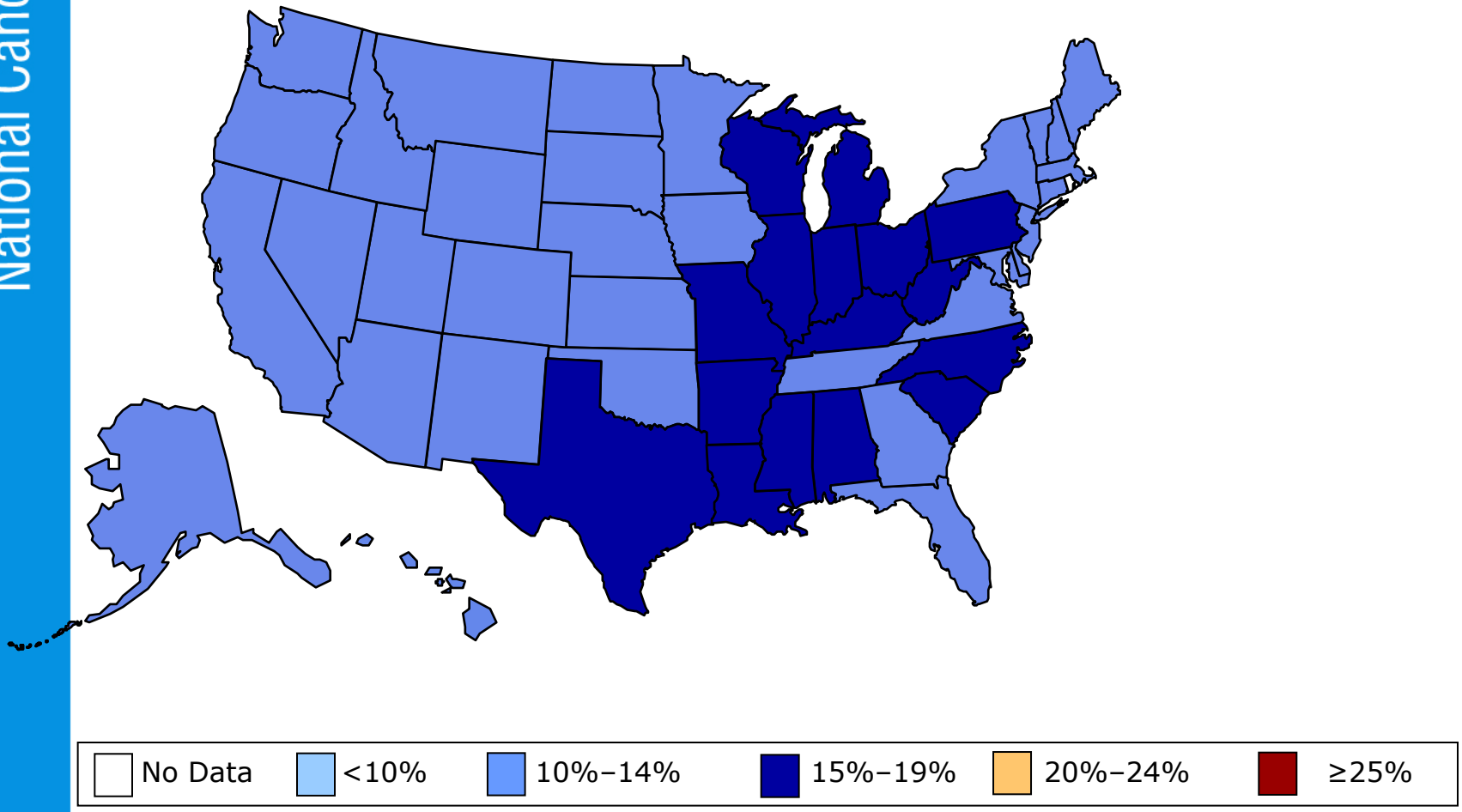


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1994

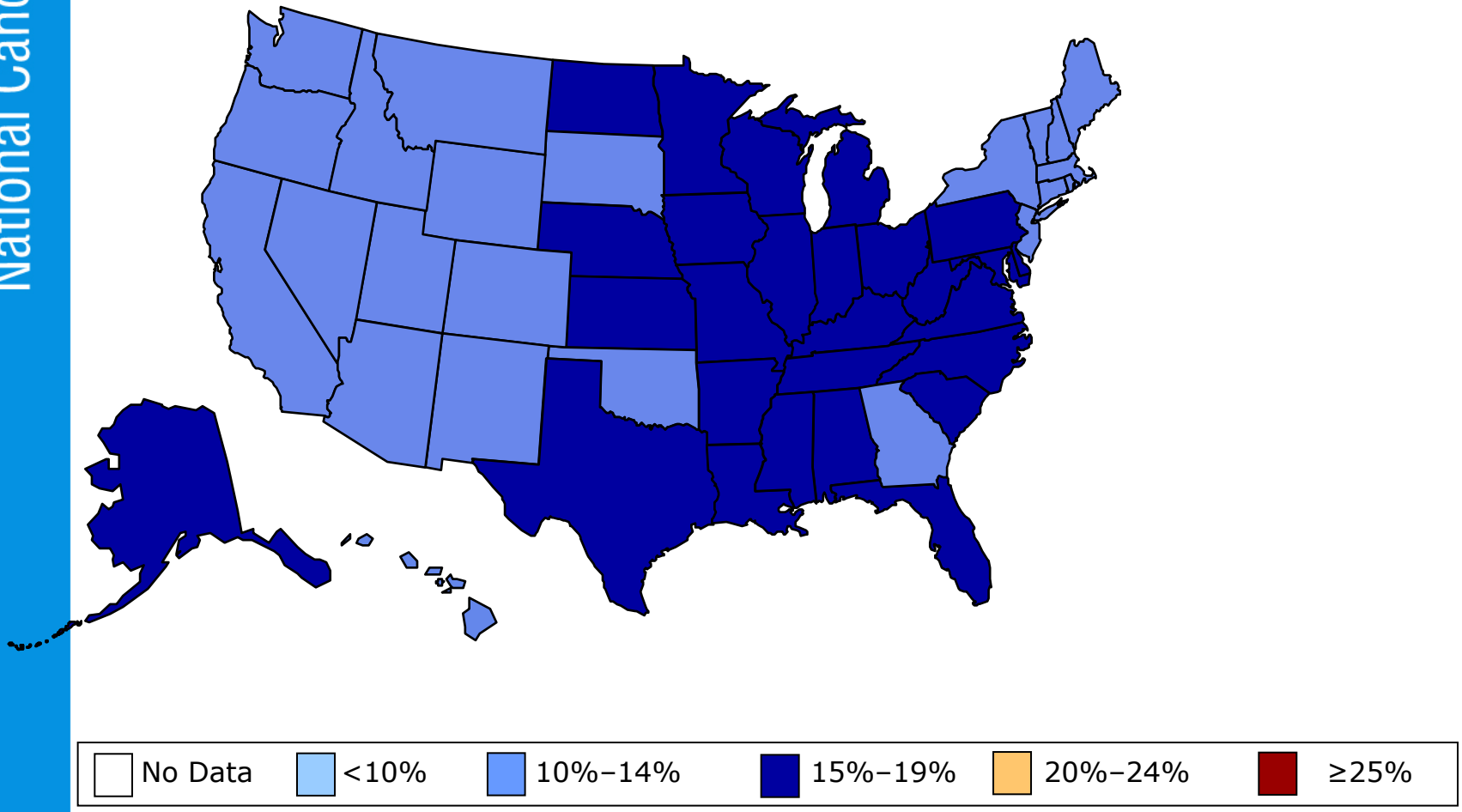


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1995

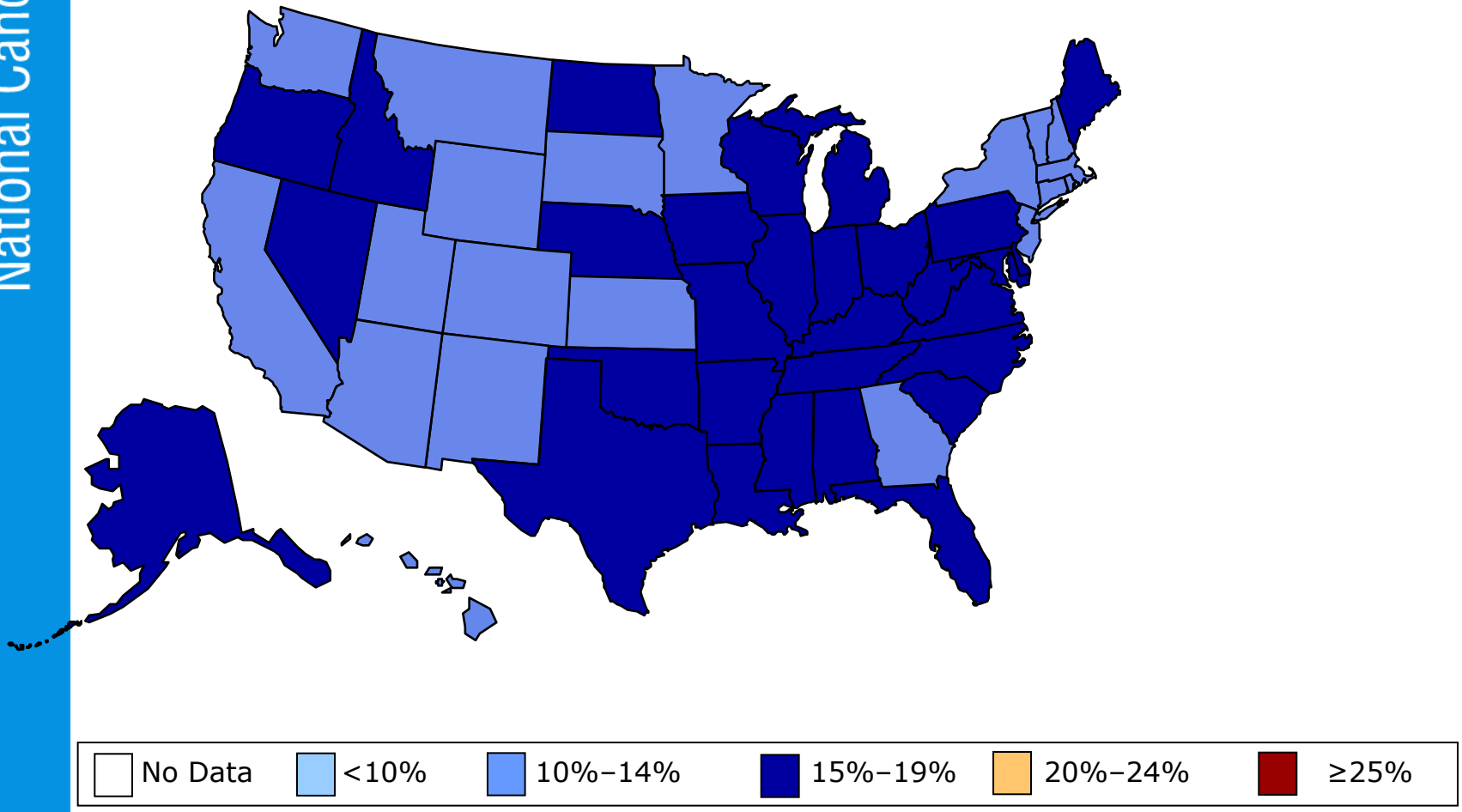


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1996

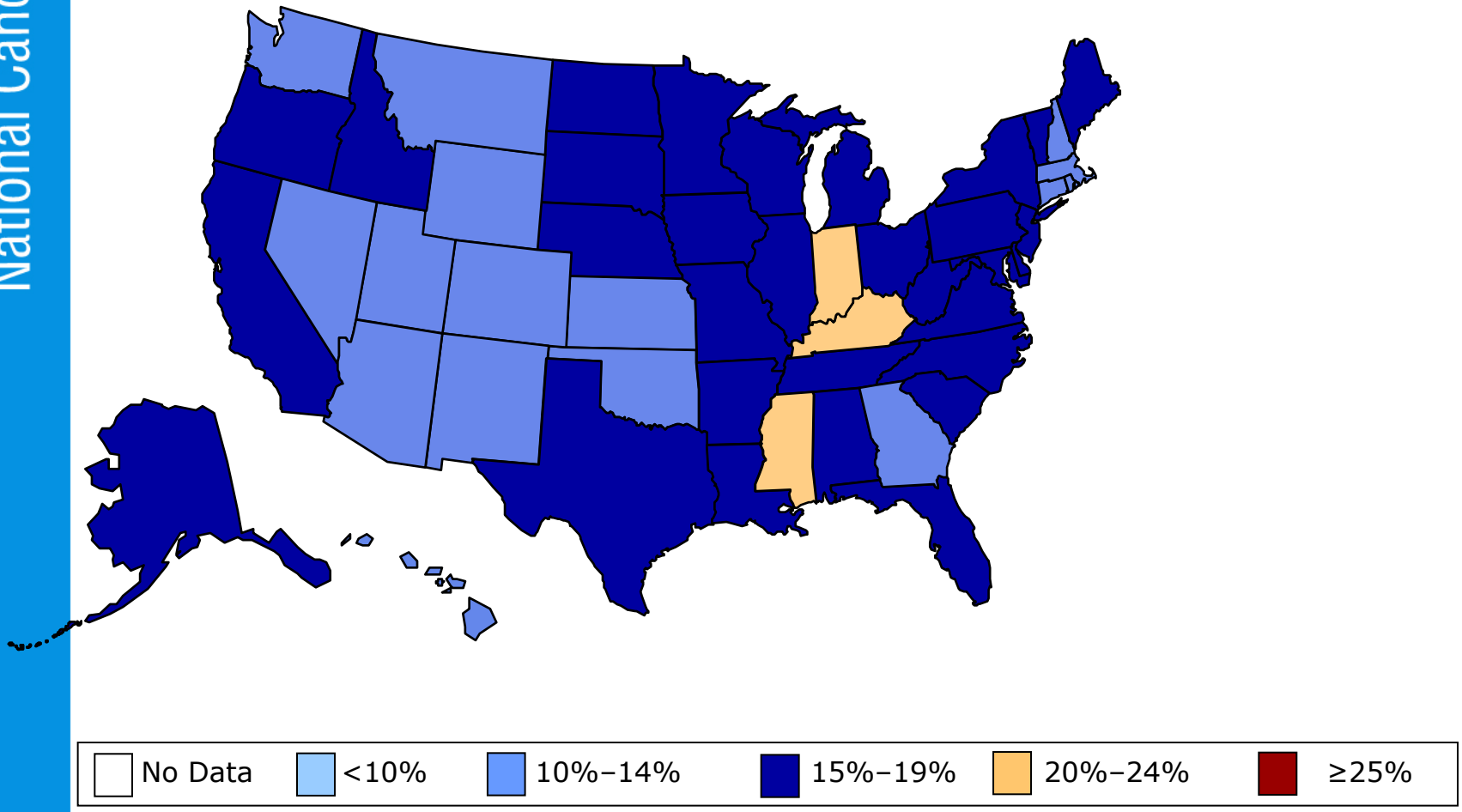


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1997

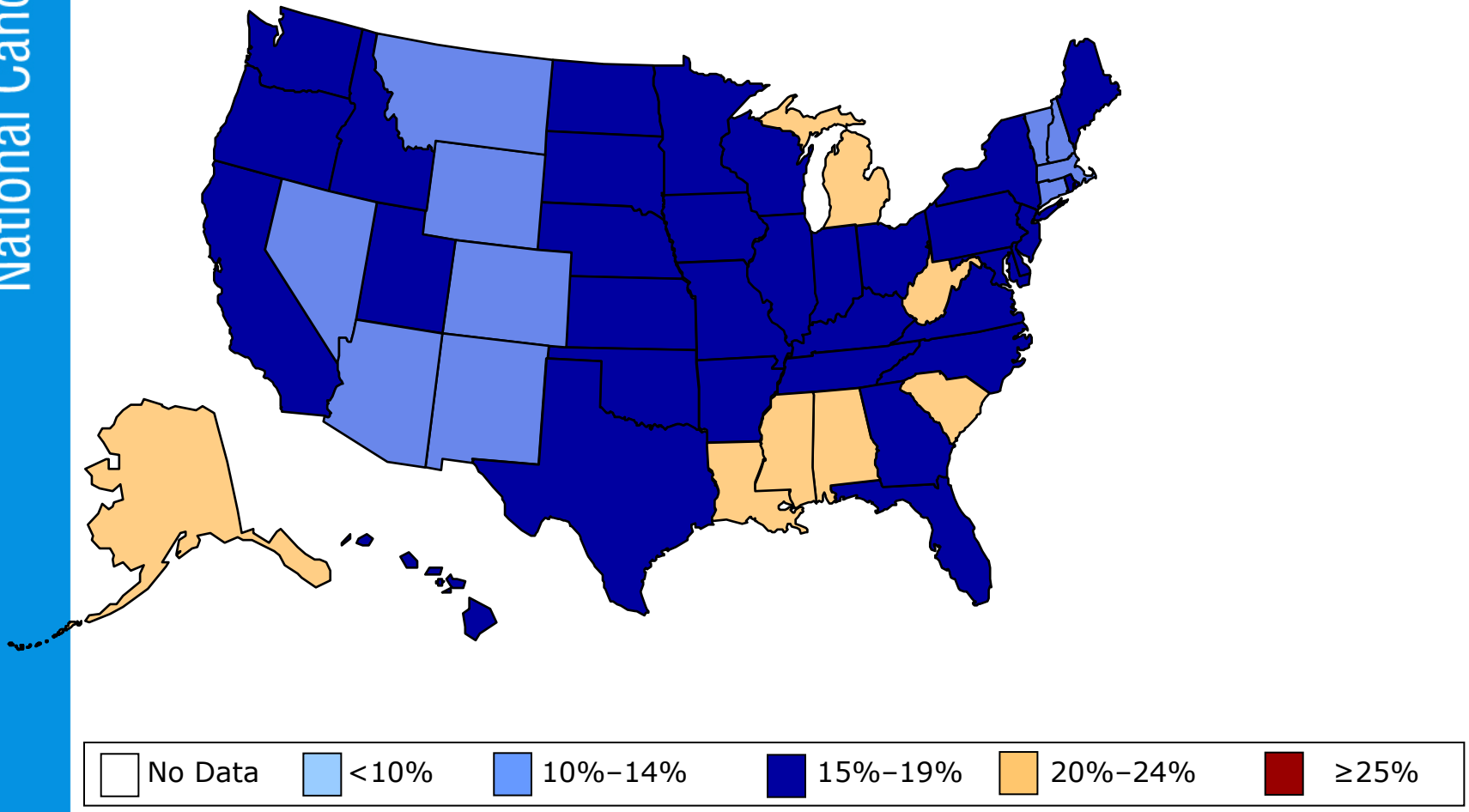


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1998

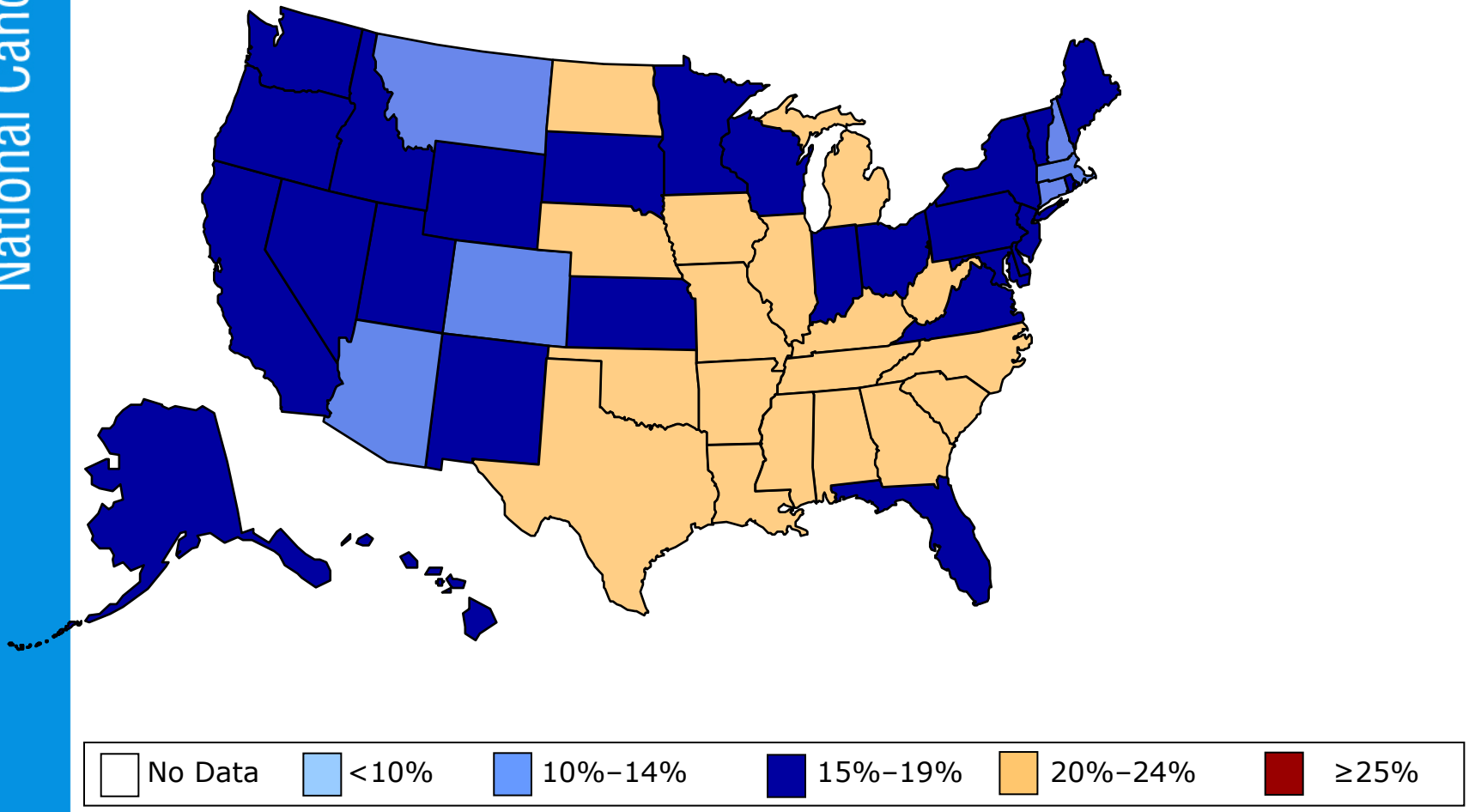


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 1999

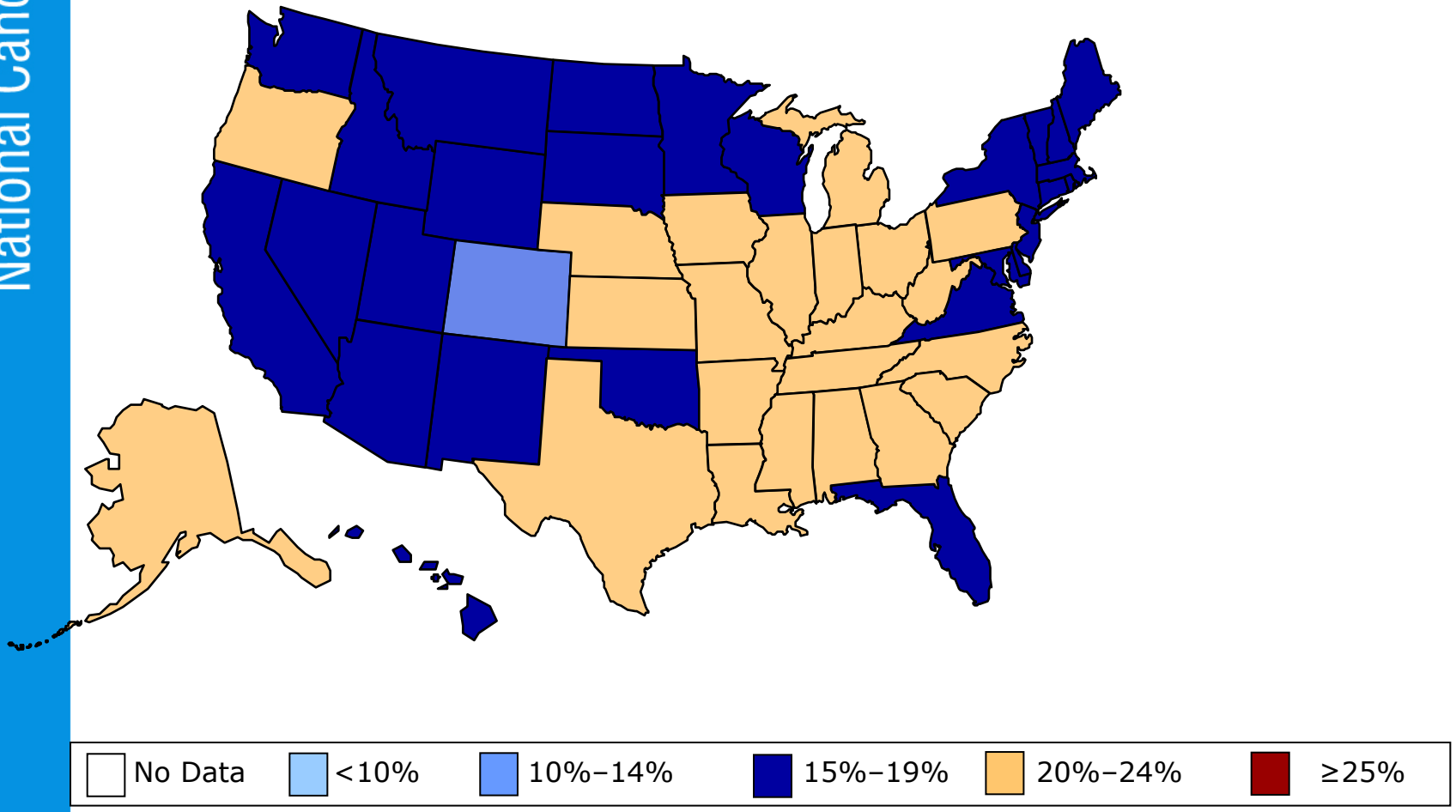


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 2000

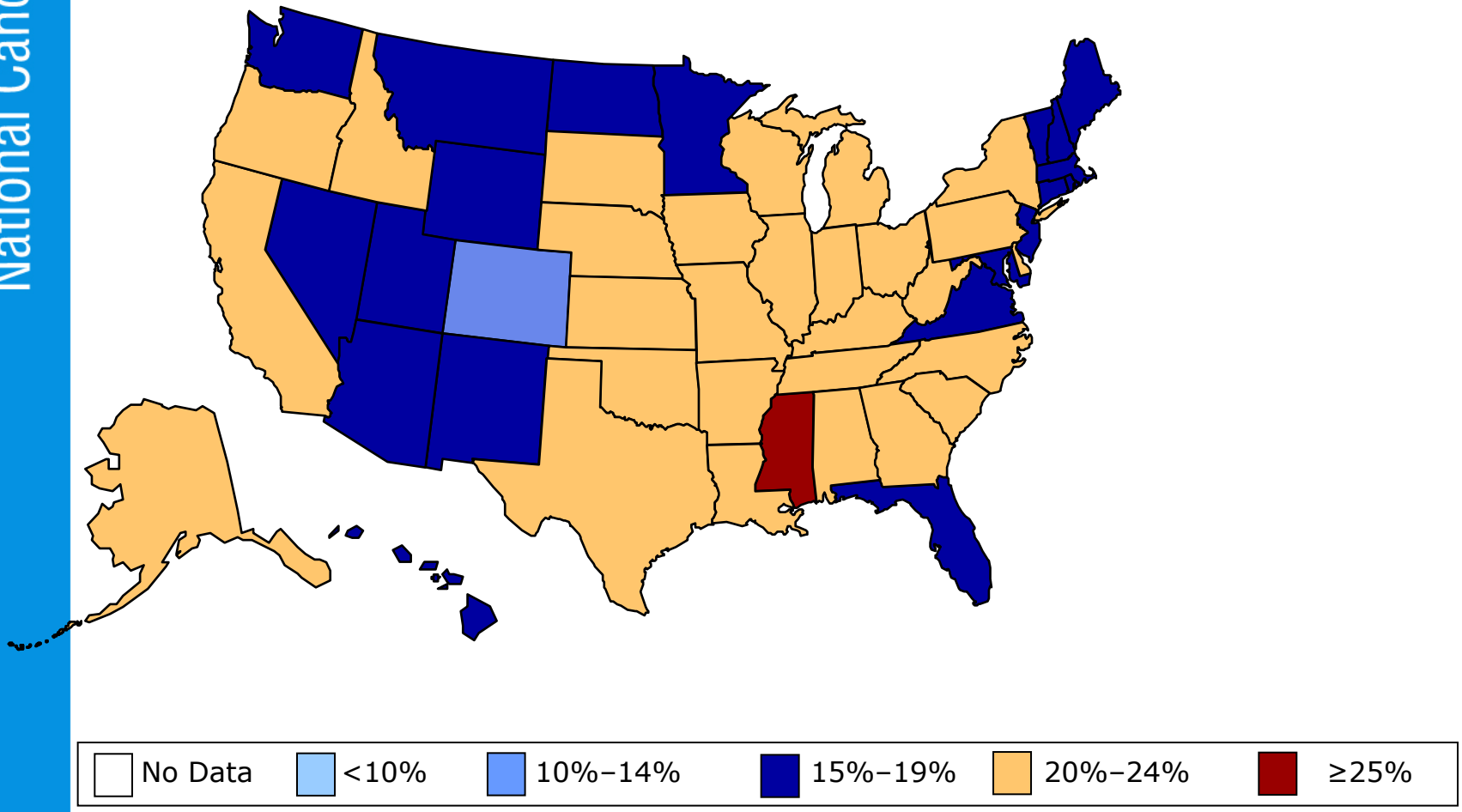


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

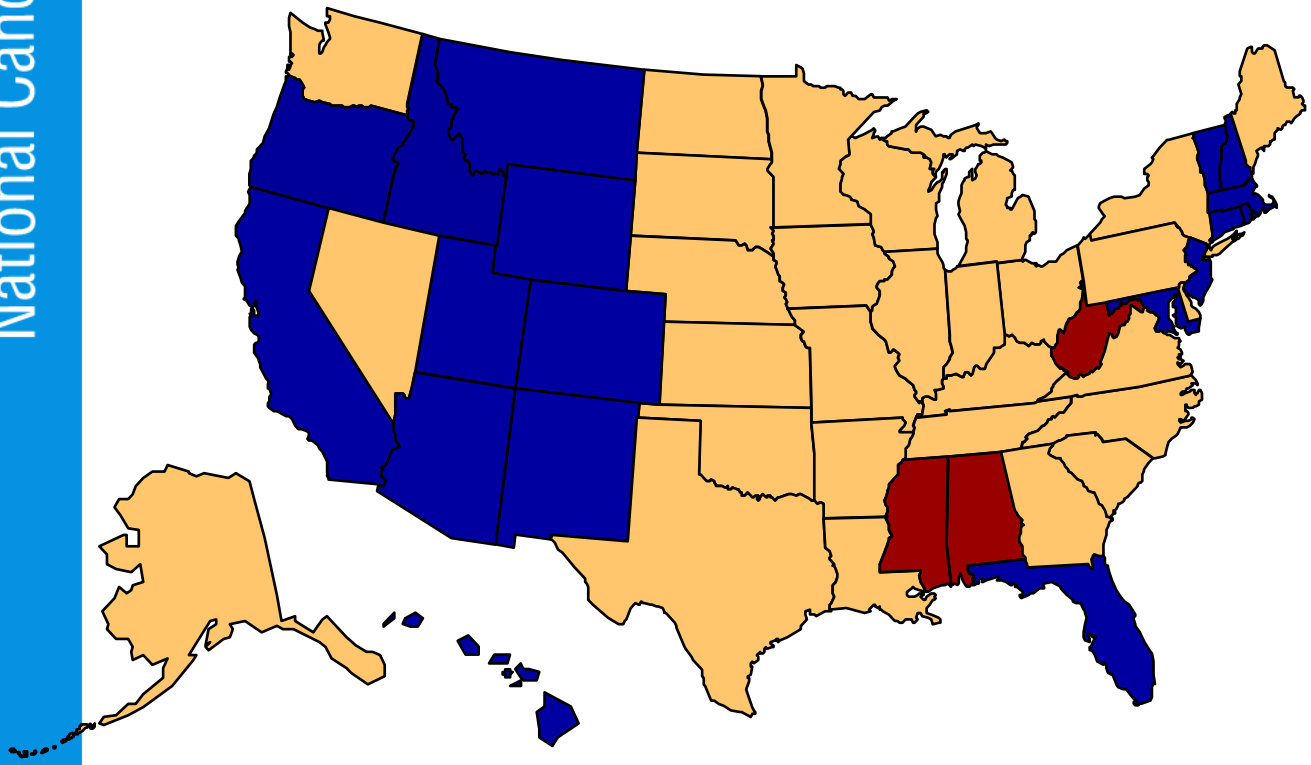
BRFSS, 2001



(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults
BRFSS, 2002

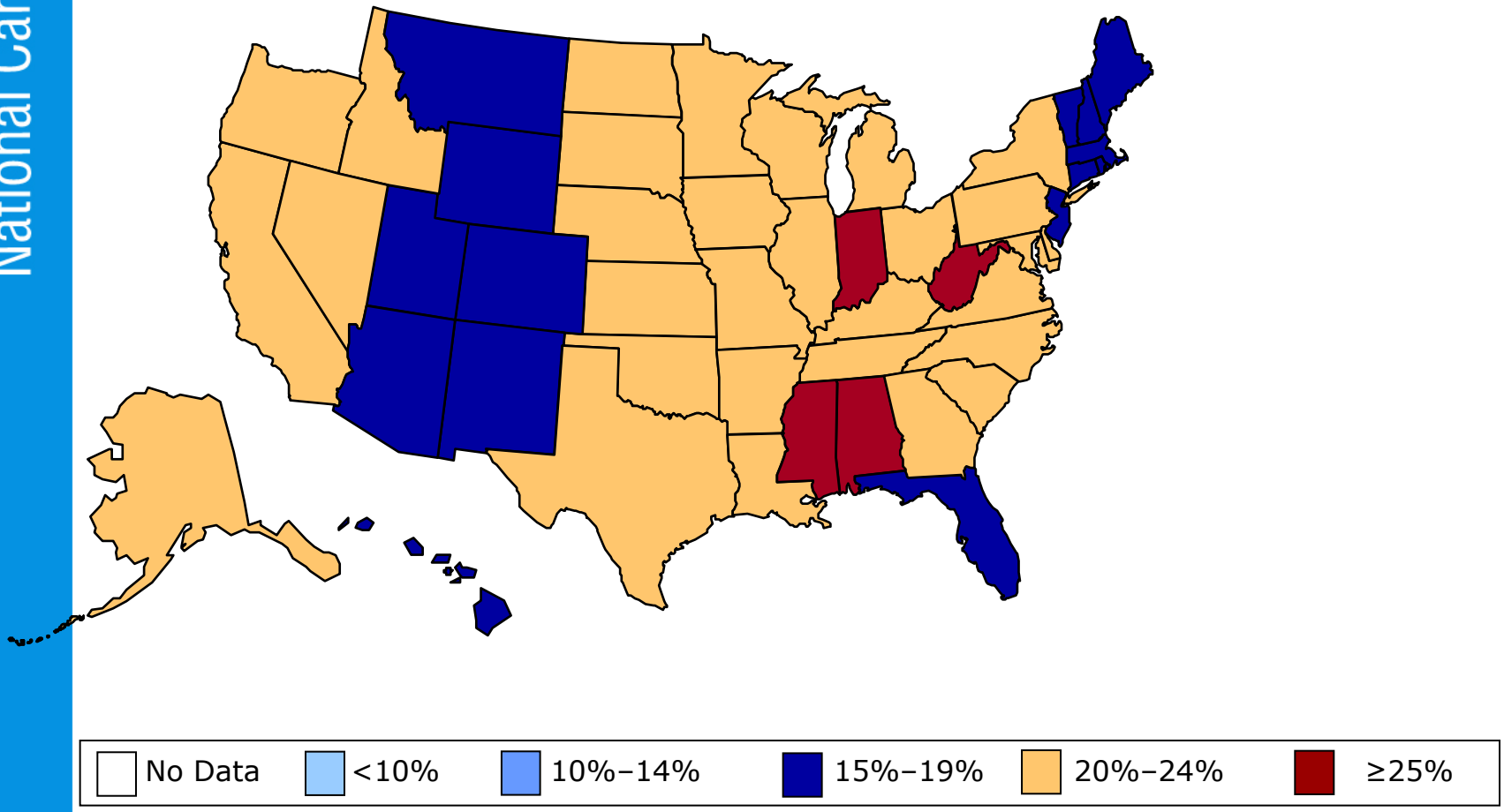


No Data	<10%	10%–14%	15%–19%	20%–24%	≥25%

(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults
BRFSS, 2003

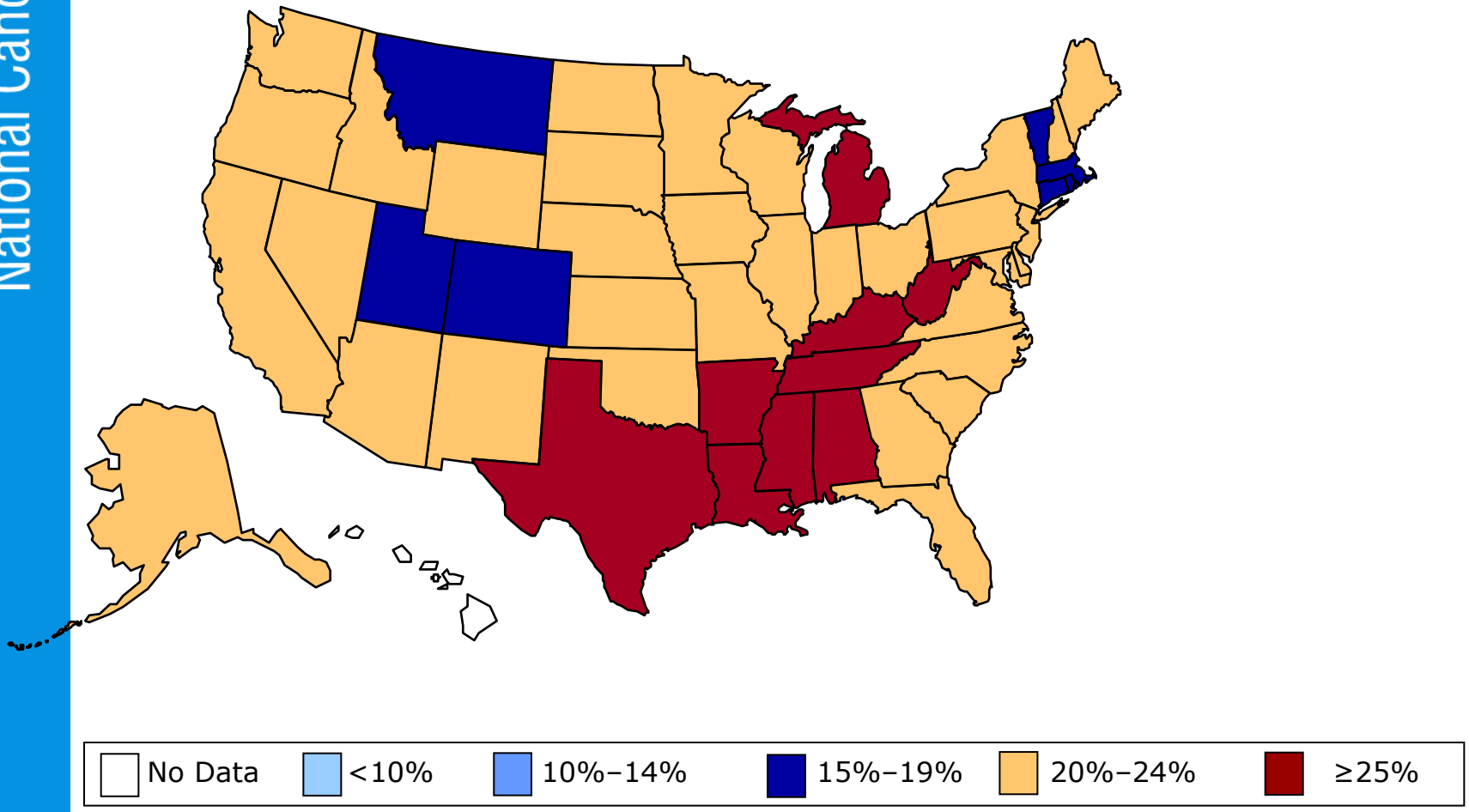


(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

Obesity Trends* Among U.S. Adults

BRFSS, 2004



(*BMI ≥30, or ~ 30 lbs overweight for 5’ 4” person)

National Cancer Institute

What do we do now and next?

- How do we change the consciousness of Epidemiologic Sciences to become more collaborative in regards to semantic interoperability???
- Propose a Populomics Health Informatics Grid, PHIG (fig) to form collaborative interest groups to develop and identify semantic ontologies and vocabularies for populations sciences that are harmonized with current efforts and to identify and make discoverable important data sets for population sciences.

National Cancer Institute

- Thank You

  - Peter Schad
  - schadp@mail.nih.gov