The Use of Non-Probability Samples to Characterize Rare Conditions

John M. Boyle, Ph.D. ICF International

Federal Committee on Statistical Methodology

November 2013

Rare Diseases

- Rare Diseases are generally defined by population prevalence, which differ somewhat across countries
- United States: a rare disease is one that affects less than 200,000 persons (1 in 1,500)
- There are an estimated 5,000 to 8,000 rare diseases
- The majority of these conditions are serious and potentially life-threatening
- An estimated 10% of the U.S. population is affected by one or more rare diseases
- In summary, rare diseases are important to public health due to the number of persons affected in the aggregate and the impact of the conditions on the affected individuals and the health care system

Evidenced Based Medicine

- Evidence based medicine (EBM) is currently treated as the best tool or gold standard to validate clinical decisions about the care of patients, individually or as a whole
- Evidence based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care off individual patients
- The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.
- Evidence based medicine is "derived from high quality research on population samples to inform clinical decision-making"
- Evidence based medicine evolved from clinical epidemiology --- using population health sciences to inform clinical practice

Evaluating Evidence in EBM

- Evidence quality can be assessed based on the source type from meta-analyses and systematic reviews of triple-blind randomized clinical trials down to conventional wisdom at the bottom
- Good quality, reliable evidence does not always have to come from clinical trials
- High quality evidence has good internal and external validity
- Critical appraisal of evidence for internal validity that can be broken down into aspects regarding:
  - Systematic errors as a result of selection bias, information bias and confounding
  - Quantitative aspects of diagnosis and treatment
  - The effect size and aspects regarding its precision
  - Clinical importance of results
 - External validity or generalizability

Primary Immune Deficiency Diseases

- Primary immune deficiency diseases (PIDD) is a class of approximately 150 specific disorders

- Immune deficiency diseases both individually and as a class of disorders are recognized as rare diseases (NIH Office of Rare Diseases)

- Library of Medicine abstract: "X-Linked Agamma-globulinemia Marks the Spot: Rare Diseases in Evidenced-Based Practice"

- Uniquely, the Immune Deficiency Foundation commissioned a national probability sample of 10,000 households sampled by random digit dialing and screened by telephone in 2005
    - Identified 18 households and 23 persons with a PIDD diagnosis
    - Estimated prevalence for the class would be 1/1,200 persons

Evidence-Based Medicine and Treatment

- EMB is usually used in the context of treatment decisions (treat or not)
- Random control trials are gold standard for assessment of efficacy and safety
- Rare diseases are a challenge for RCT in approval of new therapies
- BUT what about assessing treatment and treatment outcomes after FDA approval?
- The majority of patients with PIDD have antibody disorders for which immunoglobulin (IgG) therapy is the standard of care

Current IgG Use: IDF Patient Surveys and National Prevalence Survey
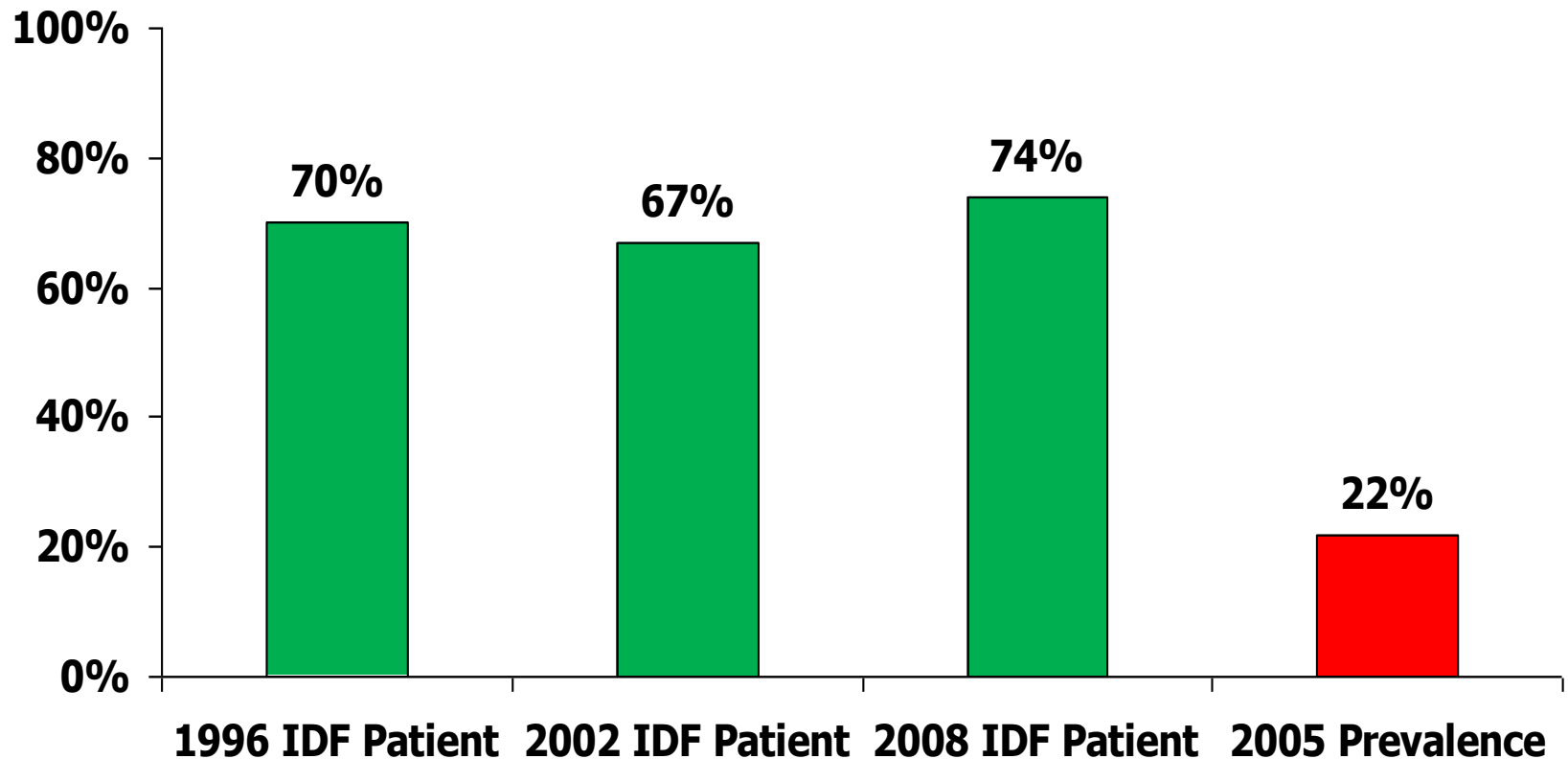
Sources: IDF Primary Immune Deficiency Diseases in America: 1996, 2002 & 2008,
2005 IDF National Prevalence Survey

2010 National PIDD Treatment Survey

- Characterize treatment with immunoglobulin replacement therapy among a national community-based sample of PIDD patients
- Obtain a sample of at least 100 diagnosed PID patients to provide more stable estimates of immunoglobulin use
    - Cost of probability survey would be prohibitive
    - Member based survey may be biased toward treatment
- Large web panel of the general public to screen for a nationally representative sample of persons diagnosed with primary immune deficiency diseases may offer a less biased, non-probability source of information on treatment

Web Survey Study Procedures

- Internet panel company sends invitations to samples of web panel members to participate in a survey
- Interested panel members go to a "requirements page" where they self-identify their eligibility for the survey
- "Eligible" respondents are given the link to the IDF survey website to conduct the treatment interview
- These respondents must answer a more detailed set of questions to determine whether they are eligible for the survey
- Respondents must have a member of their household or an immediate family member living outside of their household to be eligible for the survey
- The household or immediate family member must have a specific appropriate diagnoses of primary immune deficiency diseases
- Those respondents with eligible household or immediate family members are interviewed on-line concerning the PID patient's disease and its treatment

E-mail **Invitation** to participate in new survey opportunity   N=859,379

Went to web screener = 13%   N=114,934

Did not respond= 87%   N=744,445

Anyone in household or immediate family diagnosed with XLA, CVID, IgA def., IgG def., SCID or other immune deficiency?   Yes=3%   N=3,487

No One Qualified in Household or Family = 97%

N=111,447

Went to main questionnaire   Yes=49%   N=1,702

Did not go to Main Questionnaire

No=51%   N=1785

Has anyone in your household/immediate family ever been diagnosed with a primary immune deficiency disease? What specific types of primary immune deficiency has the (AGE) been diagnosed with?

Good PIDD Diagnosis=9%   N=144

|	
No answer to screen	10%
No PIDD in HH/Family	47%
No Diagnosis given	6%
Bad Diagnosis given	28%

PIDD Web Survey

- Goal: Obtain a sample of at least 100 diagnosed PID patients to provide more stable estimates of immunoglobulin use

- Result: Web panel yielded completed interviews with 147 eligible respondents reporting on 160 PIDD

- Result: Web panel yielded interviews with119 patients with antibody disorders for which IgG treatment is recommended

- Result: Web panel found relatively low levels of IgG treatment in a national community based sample

- Problem: How do we evaluate the quality of the evidence from a non-probability sample such as a web panel?

External Validity

- Geographic distribution --- compared to Census population distribution

- Prevalence --- compared to 2005 National Prevalence Survey (RDD)

- Treatment --- compared to 2005 National Prevalence Survey (RDD)

- Health --- compared to large, IDF member surveys (non-probability, but not convenience sample)

# External Validity: Distribution of Population and Patients by Region
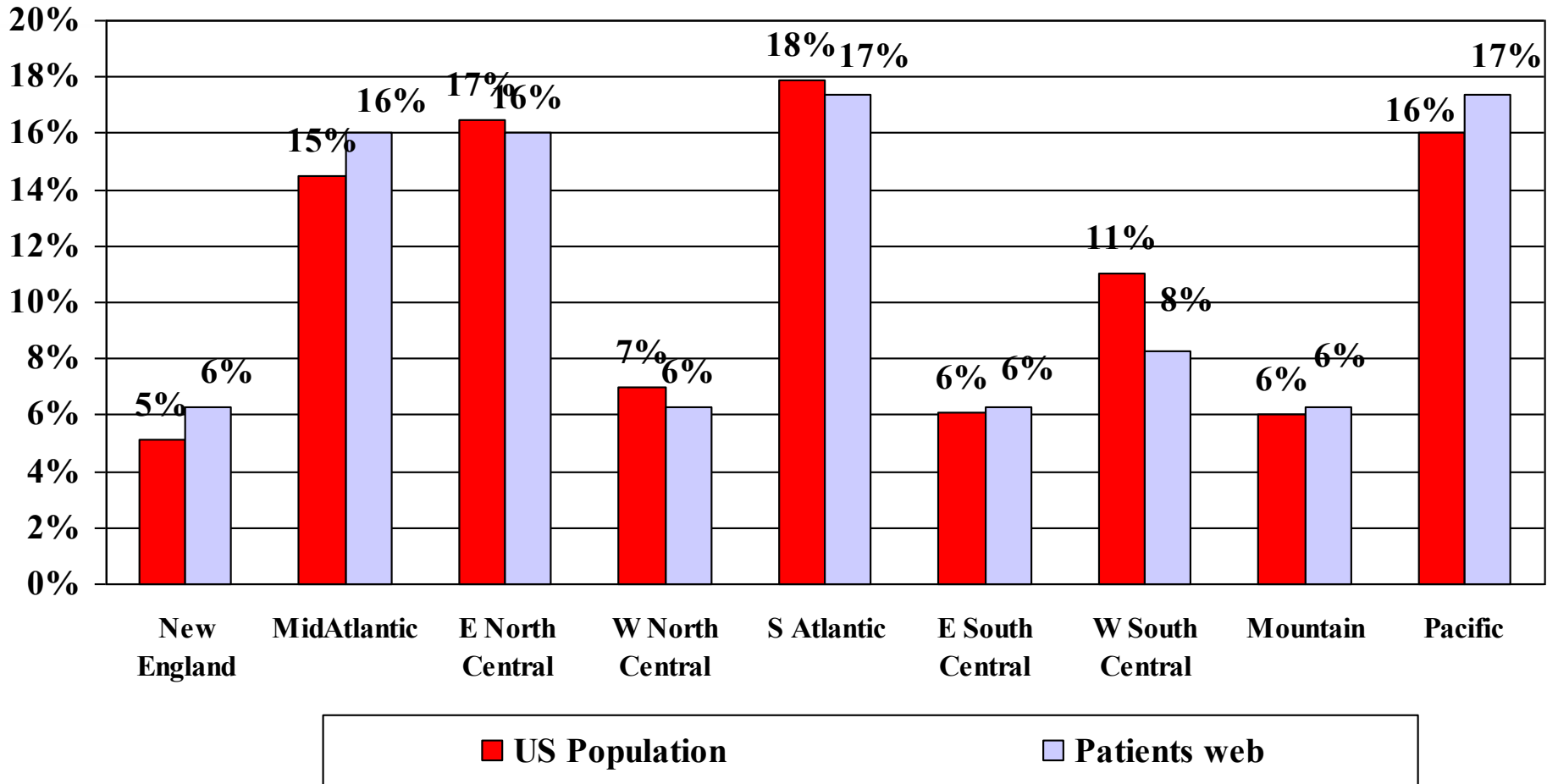
Patients in 2007 based on state of birth. Patients in web survey based on state of residence from sample file. Base: One or more eligible PIDD and completed interview N=144 (3 missing data)

External Validity:
Household Prevalence of PIDD

- 144 respondents have a qualified PIDD diagnosis in their household in web survey
- 114,934 panelists saw the screening questionnaire so
  1 out of 798 households prevalence in web survey
- If the distribution of eligible diagnoses was the same for those who qualified on the refinement question and did not go to the web survey, we would expect 295 eligible respondents for a household prevalence rate of
  1 in 390 households estimated prevalence in web survey
- Estimated household rate for diagnosed PIDD was
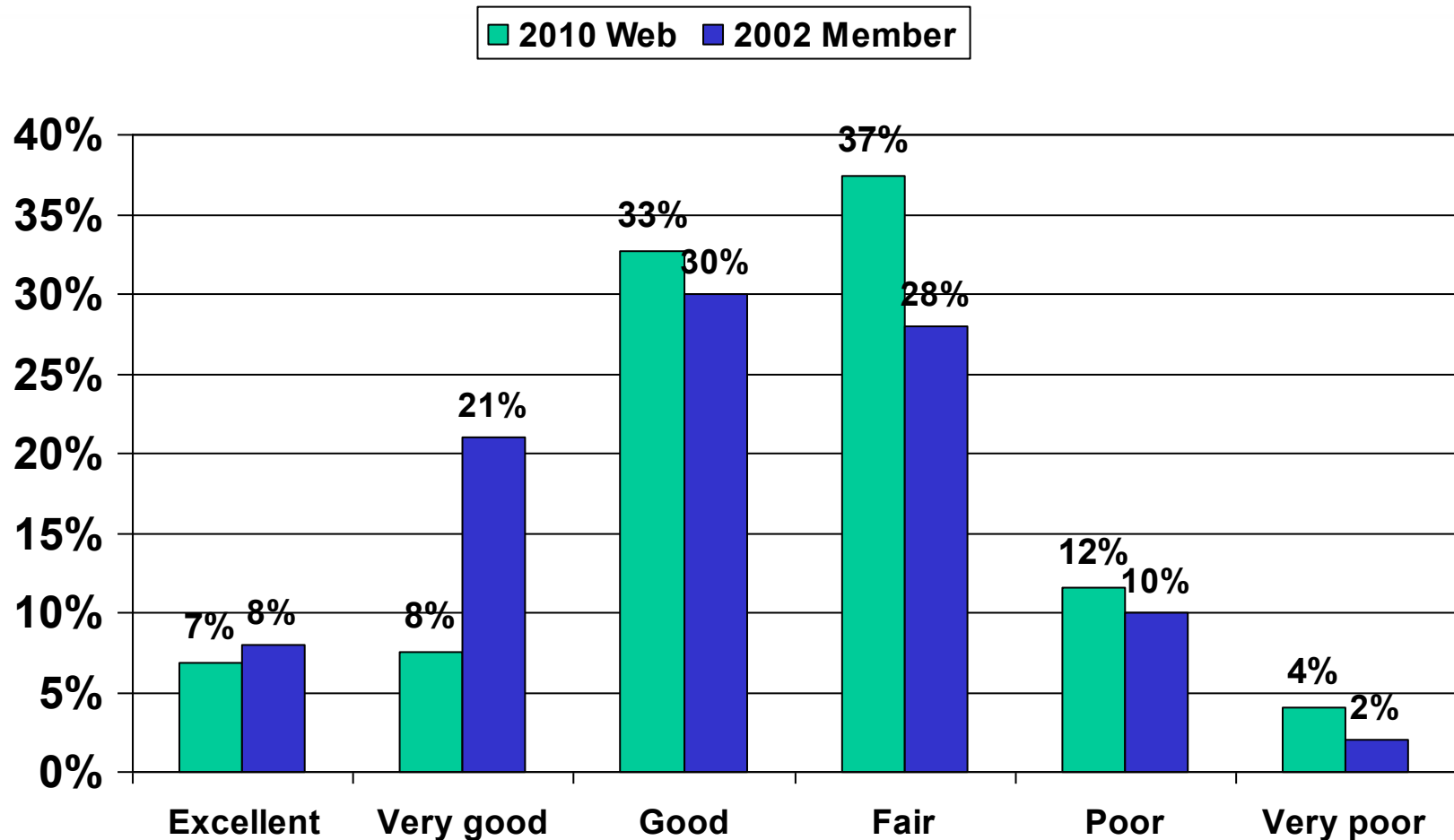 **1 in 555 households in RDD probability survey**

Compare with a Probability Survey:
Past and Current IG Treatment
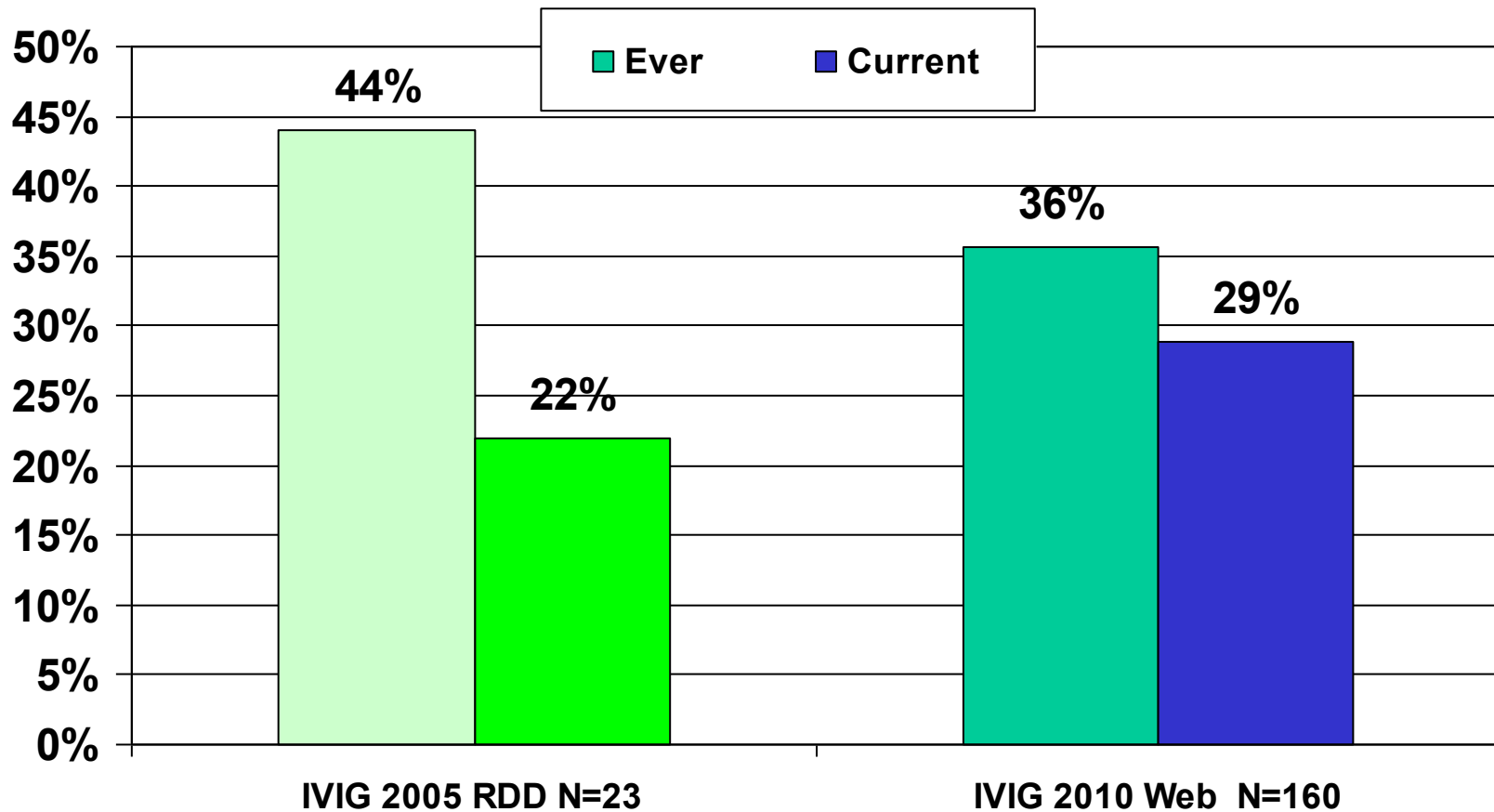
Q8a. Has the (AGE) ever been treated for a month or longer with the following?
Q8c. Is the (AGE) currently being treated with any of the following? All living eligible PIDD N=160

Compare with Other Non-Probability Survey: Current Health Status

2010 Web **2002 Member**

	Excellent	Very good	Good	Fair	Poor	Very poor
2010 Web	7%	8%	33%	37%	12%	4%
2002 Member	8%	21%	30%	28%	10%	2%

Q23. Would you describe the (AGE)'s current health status as …?
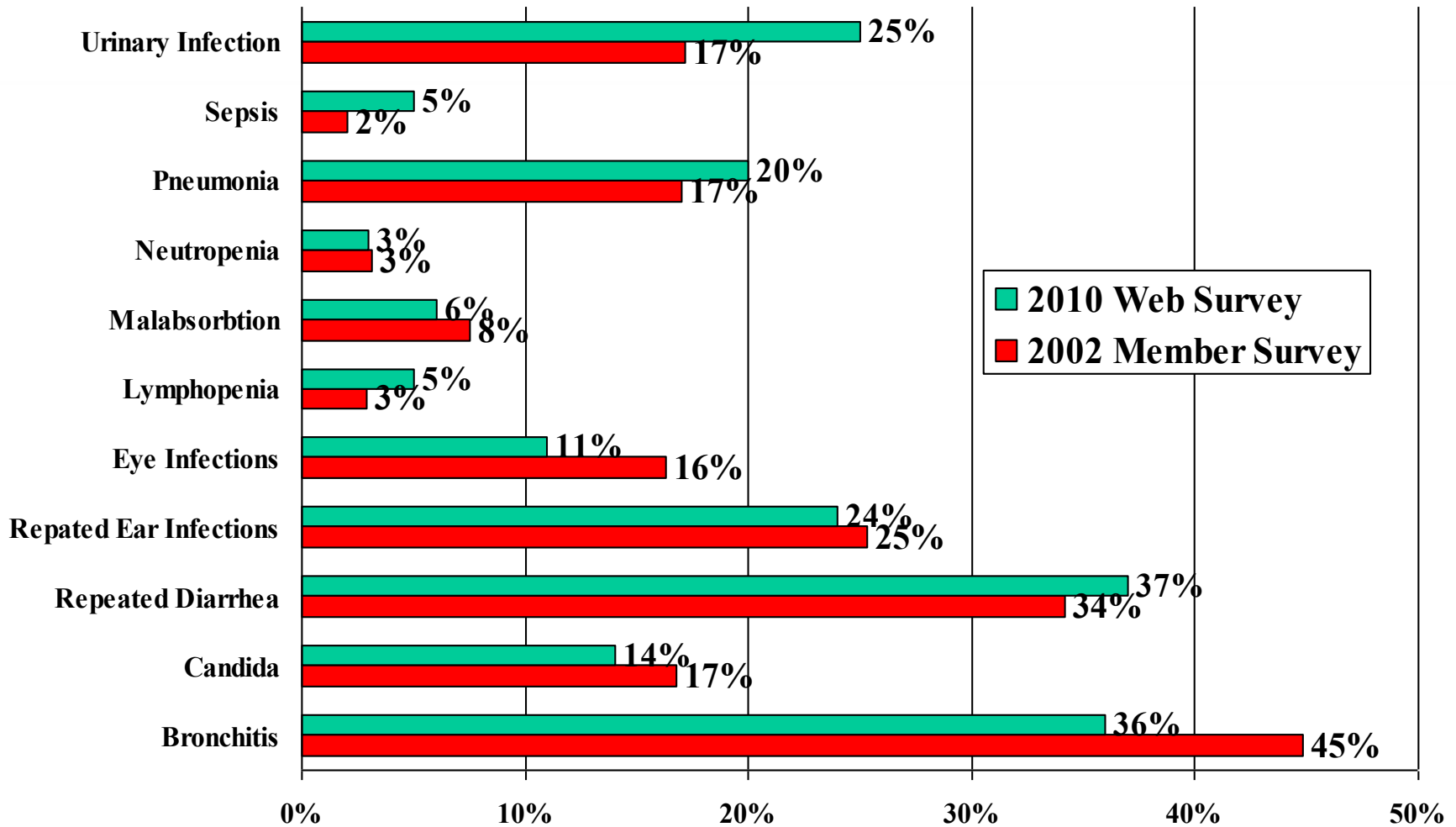Base: Selected PIDD patient N=147

Compare to Other Non-Probability Survey: Limitations due to Health
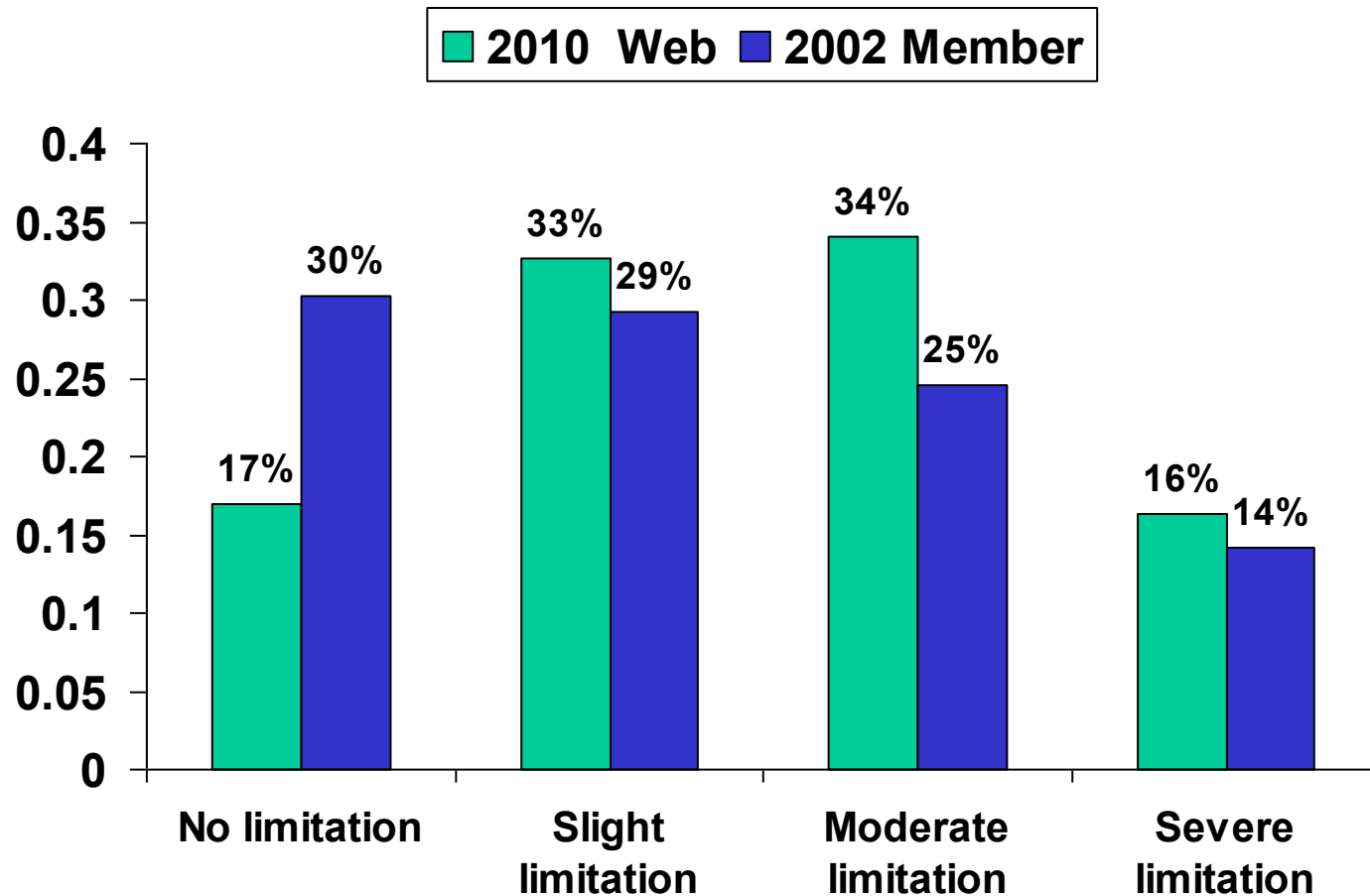
Q24. How much, if any, is the (AGE) limited in work, play or normal physical activity as a result of his/her health? Base: Selected PIDD patient   N=147

Compare with Other Non-Probability Survey: Acute Conditions in Previous 12 Months

Legend:
- ■ 2010 Web Survey
- ■ 2002 Member Survey

Condition	2010 Web Survey	2002 Member Survey
Urinary Infection	25%	17%
Sepsis	5%	2%
Pneumonia	20%	17%
Neutropenia	3%	3%
Malabsorbtion	6%	8%
Lymphopenia	5%	3%
Eye Infections	11%	16%
Repated Ear Infections	24%	25%
Repeated Diarrhea	37%	34%
Candida	14%	17%
Bronchitis	36%	45%

Q25. Which of the following conditions, if any, has the (AGE) had during the past 12 months? Base: Selected PIDD patient N=147
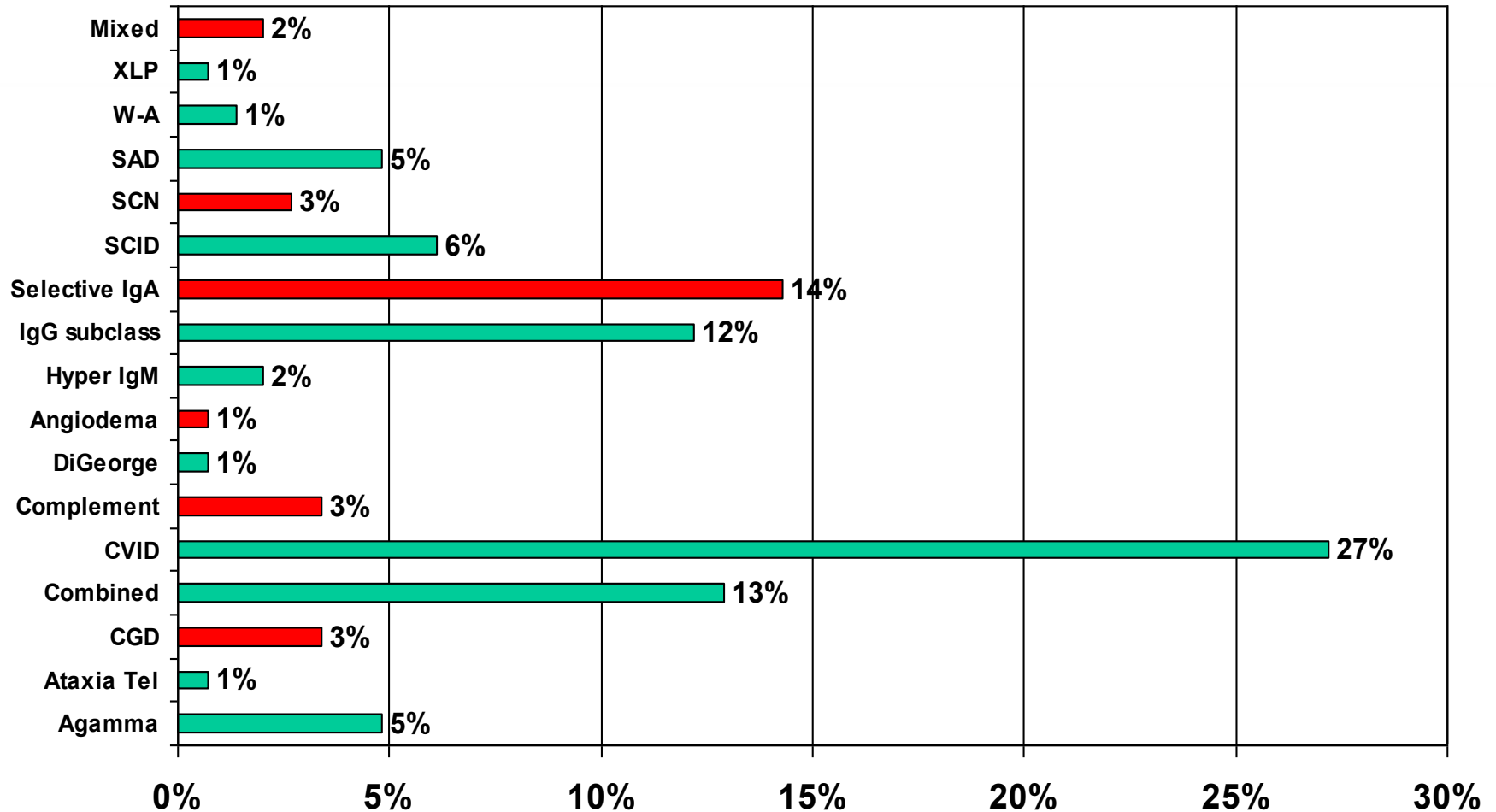
Internal Validity

- Disease specific eligibility

- Correct diagnoses by clinical expectations

- Incorrect diagnoses by type

- Location of treatment by treatment level

- Contact with patient organization by treatment level

# Disease Specific Eligibility in Web: Minimizing Fraud and Error

- Must report household or immediate family member with primary immune deficiency disease
- Must select one or more legitimate diagnosis of primary immune deficiency disease from pre-coded list of 18 legitimate and 13 non-legitimate conditions plus one "other" category on two screens
- Must report one legitimate condition, and
    - Not more than 5 persons in household (or immediate family) with PIDD conditions
    - Not three or more PIDD conditions for any person
    - Not two rare, non-PIDD conditions placed at beginning of list
    - Not combinations of PIDD conditions that were improbable

# Internal Validity: Specific PIDD Diagnoses

Q5a/b. What specific types of primary immune deficiency has the (AGE) been diagnosed with?
Base: Selected PIDD patient N=147 (Conditions using IgG in blue)

# Internal Validity: Non PIDD Diagnoses Reported as PIDD
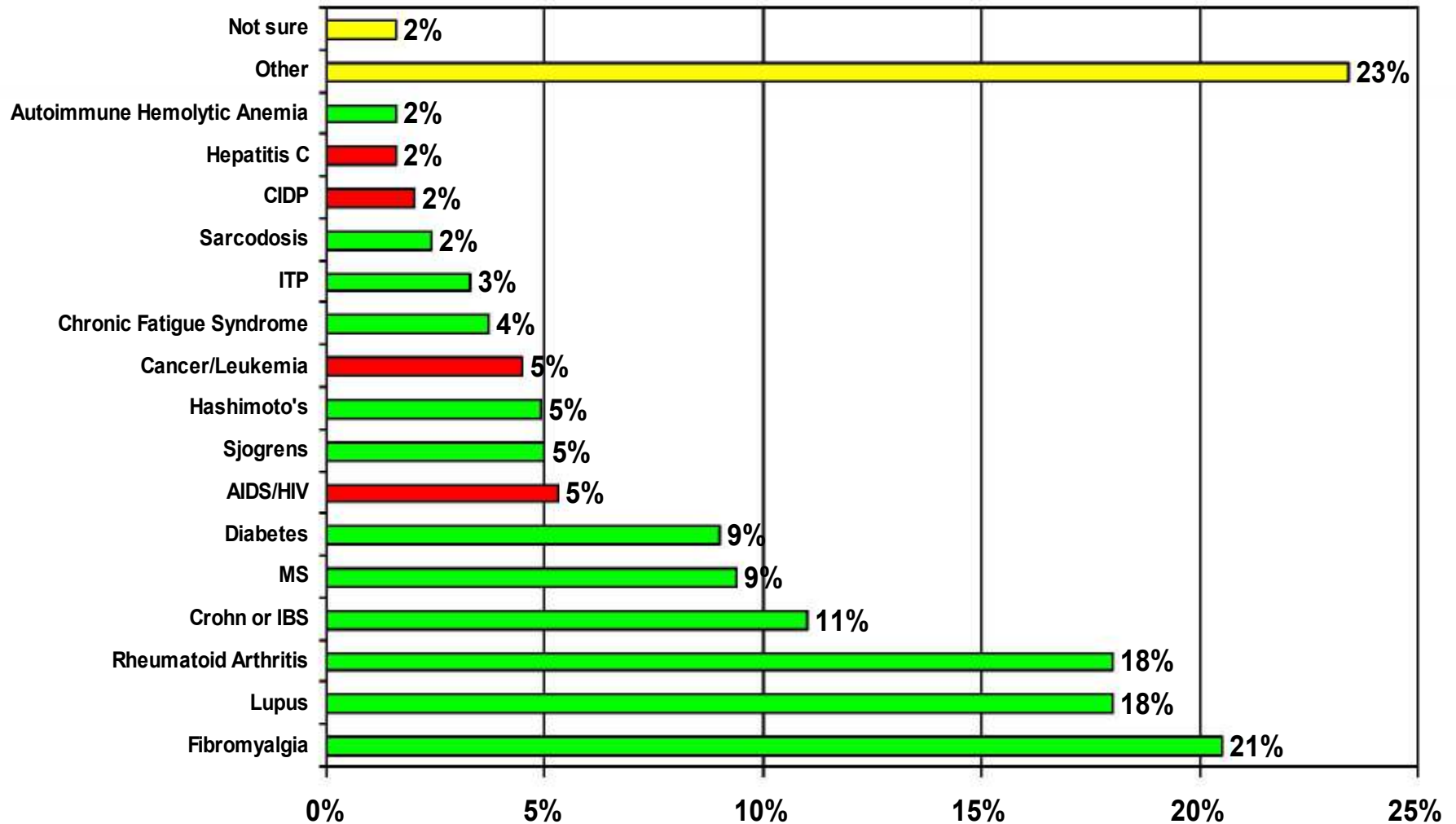
Q5a/b. What specific types of primary immune deficiency has the (AGE) been diagnosed with?
Base: PIDD Reported but Only Non-PIDD Diagnosis Given N=244 Auto-immune in Green

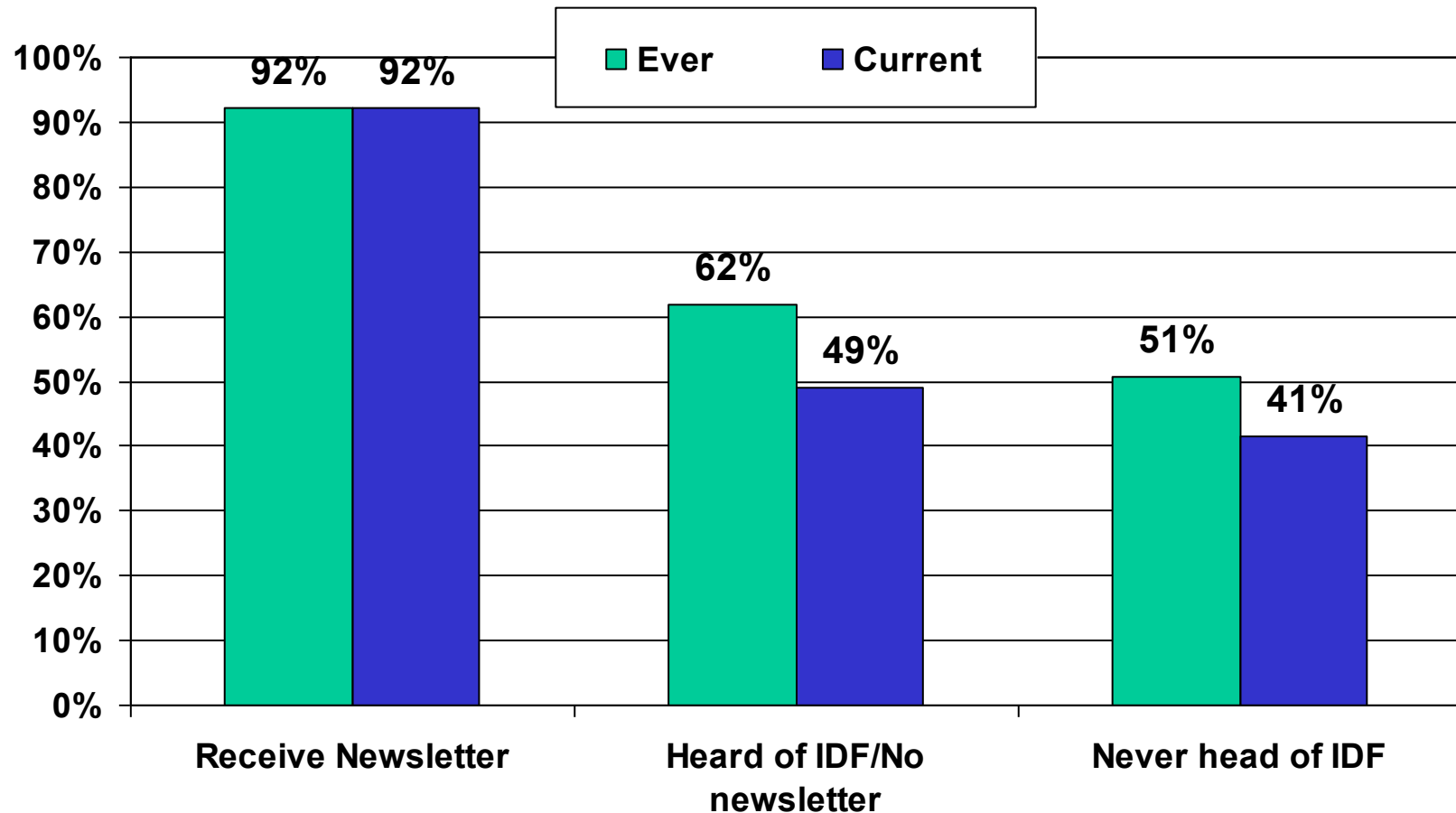# Ever and Current Treatment with Immunoglobulin by Immunologist Location
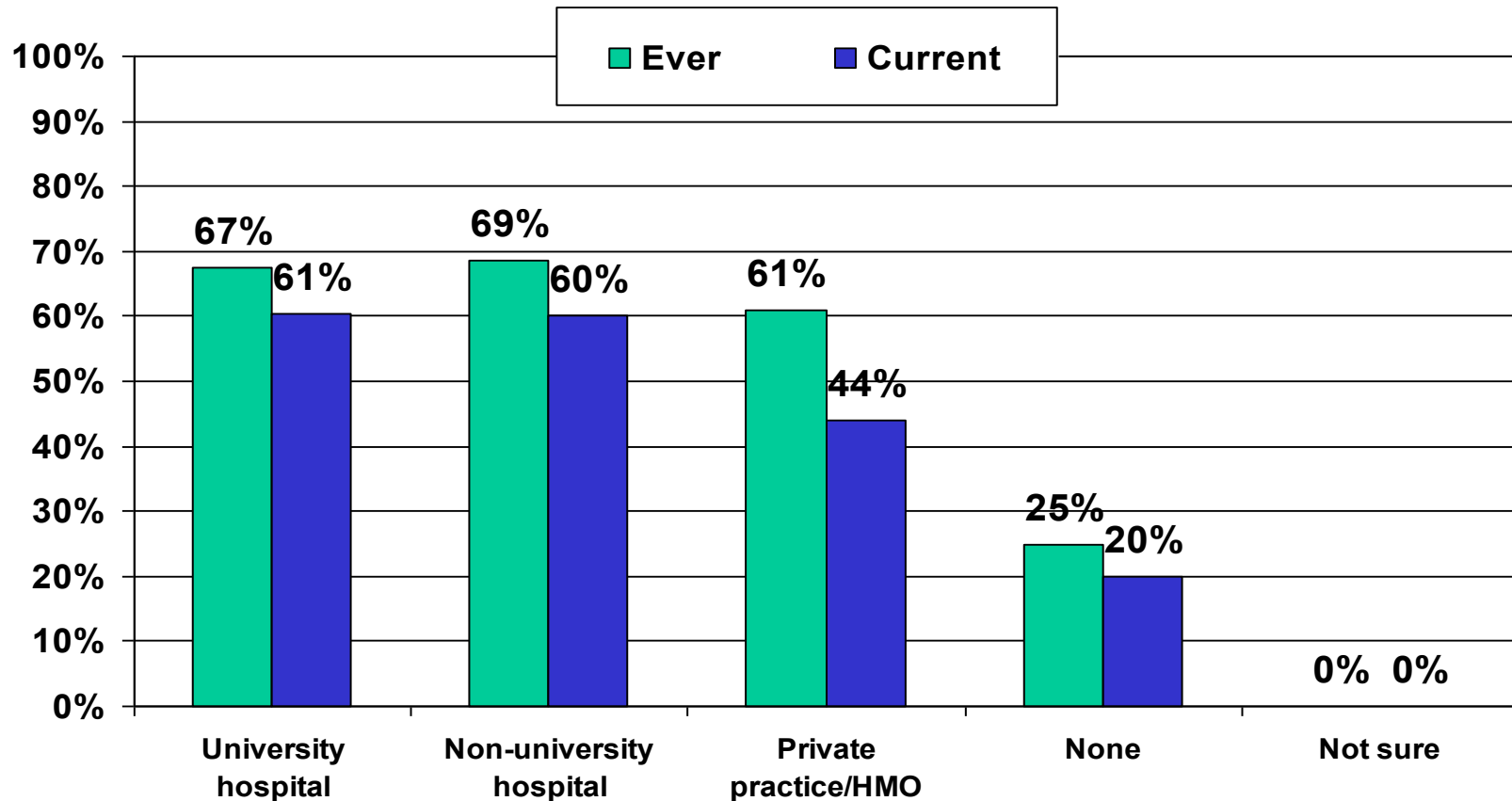
Q8a. Has the (AGE) ever been treated for a month or longer with the following? Q8c. Is the (AGE) currently being treated with any of the following? Base: Selected PIDD patient N=147

Internal Validity: Contact with IDF
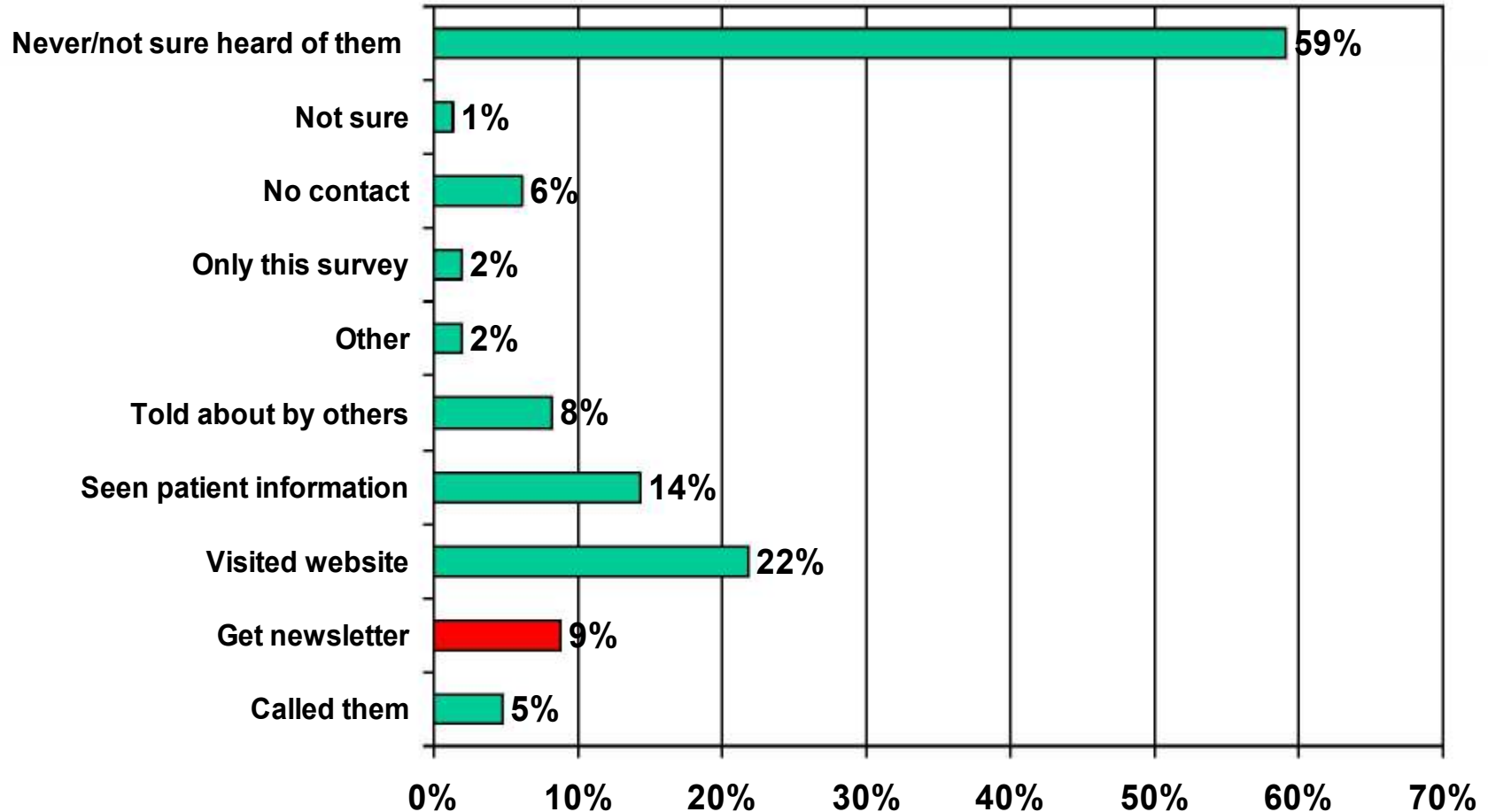
Q41b. What kinds of contact, if any, have you had with the Immune Deficiency Foundation?
Base: Selected PIDD patient N=147

Internal Validity: Treatment with Immunoglobulin by IDF Contact

Q8a. Has the (AGE) ever been treated for a month or longer with the following? Q8c. Is the (AGE) currently being treated with any of the following? Base: Selected PIDD patient N=147

Evaluating a Non-Probability Sample for Data Quality

- Patients distributed nationally similar to Census
- Household prevalence of diagnosed PIDD in web panel equivalent to RDD telephone survey
- Treatment rates similar in web and RDD surveys
- Correct diagnoses distributed to clinical expectations
- Wrong diagnoses primarily auto-immune diseases and similar to probability survey (not presented here)
- Disease characteristics of web panel patients similar to previous member surveys
- Treatment rates vary by immunologist location
- Only a minority have heard of IDF and less than one in ten would be in the IDF database
- Treatment rates of patients in web survey who receive IDF newsletter similar to IDF member surveys

Conclusions

- Incidence of rare diseases make population estimates of treatment and treatment outcomes infeasible using probability samples
- National web population panel was able to generate a national sample of a rare disease of usable size
- Web panel sample was compared to Census, probability surveys, other non-probability surveys for external validity
- Web panel sample was analyzed for internal validity for indications of fraud, error and internal consistency
- Findings from this web survey of a rare disease appear to be internally and externally consistent
- These estimates confirm a potential problem in treatment in the general community that was suggested by probability survey
- Carefully constructed and conducted surveys using an appropriate web panel can provide credible measures of the treatment of a rare disease in the community
- Can we develop a set of procedures for evaluating the quality of non-probability samples for rare diseases that would permit their use in evidence-based medicine?

Thank You

John M. Boyle, Ph.D.

Lead, Survey Research Line of Business

ICF International

john.boyle@icfi.com