Developing Performance Standards to Expedite Validation of Innovative and Improved Test Methods

Marilyn Wind[1] and William S. Stokes[2]

[1]US Consumer Product Safety Commission, Bethesda, MD, USA; [2]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA

Summary

Regulatory acceptance of scientifically valid new test methods is often followed by improved versions that incorporate innovations and enhancements to provide better performance or other advantages. The use of performance standards evolved to support the use of proprietary test methods for regulatory testing and to expedite validation of new and revised test methods that are structurally and functionally similar to accepted test methods. As new innovative technologies become available that can be used to improve existing test methods or to develop similar test methods, the availability of performance standards will facilitate more rapid evaluation of these methods. Performance standards are based on the validated reference test method and consist of essential test method components, a minimum list of reference substances, and standards for accuracy/reliability. The routine development and availability of scientifically sound performance standards is expected to expedite the efficient validation of innovative and improved test methods and testing strategies that provide for improved hazard assessments and the reduction, refinement, and replacement of animal use.

Keywords: performance standards, validation, innovative methods

1 Introduction

Performance standards communicate the basis by which new proprietary (i.e., copyrighted, trademarked, registered) and nonproprietary test methods are determined to be scientifically valid (i.e., have sufficient accuracy and reliability) for specific testing purposes. These performance standards, following acceptance by regulatory agencies, can then be used to evaluate the acceptability in terms of reliability and accuracy of other test methods that are based on similar scientific principles and measure or predict the same biological or toxic effect (Stokes et al., 2006). The development and availability of performance standards allow regulatory agencies to endorse proprietary test methods and include them in test guidelines. Performance standards also provide the basis for evaluating the acceptability of proposed test methods that are mechanistically and functionally similar to an adequately validated and accepted reference test method (Stokes and Schechtman, 2007). This paper describes the concept of performance standards and examples of recently developed performance standards for regulatory safety testing methods.

2 Elements of performance standards

Performance standards consist of three critical elements:

- *Essential test method components*
  These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed, mechanistically and functionally similar test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a proposed test method is based on the same concepts as the corresponding validated test method.
- *Minimum list of reference substances*
  These are used to assess the accuracy and reliability of a proposed, mechanistically and functionally similar test method. These substances are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method, but they should not be used to develop the decision criteria for the proposed test method. They are the *minimum* number that should be used to evaluate the performance of a proposed, mechanistically and functionally similar test method.
- *Accuracy and reliability values*
  These values provide the standards that a proposed alternative test method should meet or exceed when evaluated using the minimum list of reference substances.

3 Process for developing performance standards

ICCVAM uses a detailed, transparent process for developing performance standards for new test methods that emphasizes independent peer review and the opportunity for stakeholder involvement (ICCVAM, 2003). First, NICEATM and the

appropriate ICCVAM working group develop proposed performance standards based on available validation study data. This initial step includes input from working group liaisons designated by ECVAM, JaCVAM, and Health Canada. Alternatively, if a proposed test method sponsor proposes performance standards, these will be considered by ICCVAM at this stage.

The next step involves an international independent peer review of the proposed performance standards. A peer review panel comprised of experts from across the world evaluates the performance standards for completeness and appropriateness during its evaluation of the validation status of the proposed test method. At this stage, the proposed performance standards are also made available with the test method submission to the public for comment prior to and during the peer review panel meeting. This public review process also allows for the performance standards to be considered by other national and international advisory committees such as the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) and the ECVAM Scientific Advisory Committee (ESAC).

The appropriate ICCVAM working group, with the assistance of NICEATM, prepares the final performance standards for ICCVAM approval, taking into consideration all of the recommendations of the peer review panel, advisory committees, and public comments. Performance standards recommended by ICCVAM are provided to Federal agencies and made available to the public so that they then can be referenced as adopted by regulatory authorities in guidelines issued for new test methods.

4 Critical step: developing the reference substance list

The process for developing a list of performance standards references substances should follow a step-wise process and consider several selection criteria during this process. An initial list of candidate reference substances can be generated simply by identifying those within the validation database that are 1) commercially available and 2) have high quality reference data demonstrating consistent results in the validated reference test method and in the *in vivo* or other reference test method. Substances that do not meet these criteria should not normally be considered as a reference substance. Once a qualifying list has been developed based on these essential criteria, additional selection criteria can be applied to prioritize the reference list. To the extent possible, the final list of performance standards substances should:

- Represent the full range of responses that the validated test method is capable of measuring or predicting, from strong to moderate to weak effects, as well as negative effects
- Represent the relevant range of chemistry and chemical classes
- Representative the relevant range of physical properties (e.g., solids/liquids, molecular weight, solubility)
- Have data or experience (e.g. accidental exposures) available from the species of interest; i.e., human data if proposed for predicting human effects
- Reflect the accuracy of the validated test method
- Have well-defined chemical structures
- Not be associated with excessive occupational or environmental hazard or prohibitive disposal costs
- Be readily available from commercial sources
- Have high quality data available from the validated reference test method and the *in vivo* or other reference test method

5 Development of performance standards for *in vitro* corrosivity test methods

Following the concepts detailed above, ICCVAM developed performance standards for three proprietary dermal corrosivity test methods previously recommended by ICCVAM: Corrositex®, EPISKIN™, and EpiDerm™ (ICCVAM, 1999, 2002). ICCVAM also developed performance standards for the one non-proprietary test method (the rat skin transcutaneous electrical resistance [TER] assay) that was also recommended by ICCVAM (ICCVAM, 2002). Due to the structural and functional differences of these test methods, three different sets of performance standards were developed (ICCVAM, 2004).

One set of performance standards was based on the reconstructed human skin model systems (i.e., EPISKIN™, and EpiDerm™). Because the validation database was larger for EPISKIN™ at the time, the standards were based on that method. In addition to the essential test method components, a minimum list of 24 substances was selected from the 60 substances used for validation of EPISKIN™; this included 12 corrosives and 12 noncorrosives. The decision criteria for these assays (as well as the rat skin TER) do not allow detection of all United Nations corrosivity packing groups and instead distinguish between Category I, a combined Category II/III, and not corrosive. As a result, fewer substances were required for a balanced design and the final minimum substances included 12 corrosive substances and 12 noncorrosive substances. The 24 substances were selected based on commercial availability, representation of the full severity range of dermal corrosivity, and representation of relevant chemical classes.

Another set of performance standards were developed for the rat skin TER, in which a minimum list of 24 substances was selected from those in the total validation database of 60 chemicals. Again, commercial availability, representation of the full severity range of dermal corrosivity, and representation of relevant chemical classes were used as criteria for selection. NICEATM and ICCVAM recently submitted proposals to OECD to update the TGs for these test methods (TG 430 [rat skin TER] and TG 431 [human skin model systems]) with performance standards (OECD, 2009a, 2009b).

The third set of performance standards were developed for membrane barrier systems like Corrositex®. The candidate list used to select the proposed minimum reference substances for the membrane barrier was initially generated from the original validation database of 163 substances. This was reduced

to 40 substances after considering the commercial availability of substances, representation of severity range of dermal corrosivity (UN Packing Groups 1, 2, and 3); and representation from relevant chemical classes. In order to allow detection of all 4 United Nations corrosivity packing group categories (I, II, III, and not corrosive), the final list needed sufficient numbers of a balanced design. This resulted in the final selection of 12 Noncorrosive substances, 9 Packing Group I substances, 9 Packing Group II substances, and 10 Packing Group III substances. OECD Test Guideline (TG) 435 (Membrane Barrier Systems) is the first OECD TG to include test method performance standards (OECD, 2006).

6 Putting concepts into practice: performance standards for the murine Local Lymph Node Assay (LLNA)

Internationally harmonized performance standards for the LLNA were recently developed through collaboration between NICEATM-ICCVAM, ECVAM, and JaCVAM. These performance standards will allow rapid assessment of the validity of modified versions of the traditional LLNA, such as those using non-radioisotopic methods.

The candidate list used to select the proposed minimum reference substances for the LLNA performance standards was initially generated from the database submitted to ICCVAM for the 1998 evaluation of the LLNA. This database of 209 substances was reduced to 127 candidate substances by identifying those substances for which comparative guinea pig maximization (GPMT) or Buehler test (BT) data were available. The availability of such data is important because any accuracy comparisons of new or revised methods must include the currently accepted regulatory test methods (i.e., in this case, the LLNA, and the GPMT and/or BT), as well as comparison to available human data and/or experience. Limiting the list to substances with GPMT and BT data that were collected using a standard protocol (e.g. EPA, 1998) and those with unequivocal LLNA results reduced the set from 127 to 97. Substances must also be readily available from commercial sources. Further limiting the list of substances to those that are commercially available reduced the list from 97 to 81 candidate substances.

The candidate list was then reduced to a candidate list of 40 substances taking into consideration, where feasible, the following criteria:

– Maintaining similar accuracy statistics to those achieved in the original LLNA validation report
– Availability of human testing data or experience

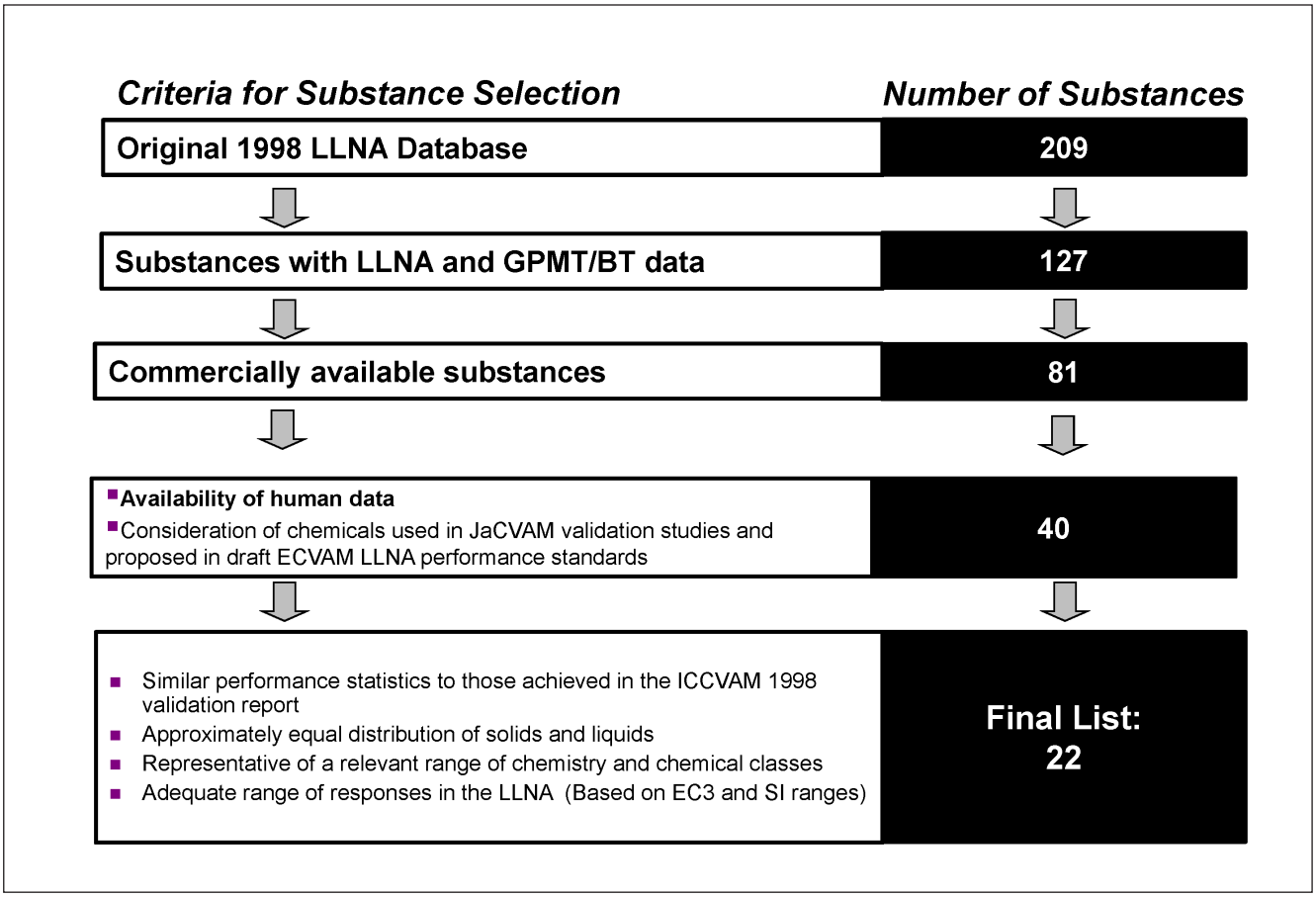

Fig. 1: Process for developing a minimum list of LLNA performance standards substances.

Tab. 1: Timeline for the Development of ICCVAM Performance Standards for the LLNA

Date	Event
Jan 10, 2007	ICCVAM nomination from the CPSC
May 17, 2007	ICCVAM public request for comments, Panel nominations, and data (NIEHS, 2007a)
Sep 12, 2007	Draft ICCVAM LLNA Performance Standards (PS) released for public comment (NIEHS, 2007b)
Sep 25-27, 2007	ECVAM Workshop on an evaluation of performance standards and alternative endpoints for the LLNA (Basketter et al., 2008).
Jan 8, 2008	Revised draft PS published for public comment (NIEHS, 2008)
Mar 4-6, 2008	ICCVAM Independent Peer Review Panel Meeting, CPSC Headquarters, Bethesda, MD; public meeting with opportunity for oral public comments.
Sep 23-24, 2008	ECVAM/JaCVAM/ICCVAM meeting on internationally harmonized LLNA PS
Oct-Nov 2008	ICCVAM (Oct 29) and ESAC (Nov 5) endorse harmonized LLNA PS
Jun 2009	Circulation of revised OECD TG 429 with LLNA PS

– Maintaining approximately the same proportion of solids and liquids
– Representing a relevant range of chemistry and chemical classes
– Providing an adequate range of responses in the LLNA, from strong to weak to negative
– Consideration of substances used in the Japanese Center for the Validation of Alternative Methods (JaCVAM) validation studies and in draft performance standards proposed by ECVAM

A final list of 22 proposed reference substances was then selected from the list of 40 candidate substances based on the selection criteria (ICCVAM, 2009). Figure 1 provides a breakdown of the impact of specific criteria on the list of candidate substances.

7 Development of internationally harmonized LLNA performance standards

ICCVAM released draft performance standards to the public for comment on September 12, 2007 (NIEHS, 2007b). Concurrently, two other international validation organizations, ECVAM and JaCVAM, were also independently developing LLNA performance standards. ECVAM was independently drafting LLNA performance standards that could be used to evaluate a non-radioactive LLNA test method submitted from a European developer (Ehling et al., 2005), and JaCVAM was drafting performance standards that could be applied to two non-radioactive LLNA methods for which validation studies were underway (Takeyoshi et al., 2001; Idehara et al., 2008).

Harmonized performance standards were viewed as critical for the success of efforts to reduce, refine, and replace the use of animals in regulatory testing for allergic contact dermatitis testing. Therefore, NICEATM and ICCVAM invited ECVAM and JaCVAM to designate liaisons to the ICCVAM Immunotoxicity Working Group (IWG) in order to work closely together to develop internationally harmonized performance standards. Input was also obtained from the ECVAM Task Force on Skin Sensitization (Basketter et al., 2008).

After consideration of these comments, a revised version was made available to an ICCVAM Independent Expert Peer Review Panel (Panel) for consideration at a public meeting in March 2008 (ICCVAM, 2009). The revised draft performance standards were also made available to the public for comment in advance of the Panel meeting, and all comments received were provided to the Panel for their consideration. The Panel's conclusions and recommendations were made available to the public and to SACATM and ESAC for comment. The Panel Report and all comments by the public, ESAC, and SACATM were considered by ICCVAM in preparing final LLNA performance standards recommendations for public release and submittal to U.S. Federal agencies (ICCVAM, 2009). Table 1 provides a summary of the timeline associated with the development of harmonized LLNA performance standards, which recently culminated with a proposal to update OECD TG 429 (OECD, 2002) with these performance standards.

8 Conclusions

Performance standards allow for expedited validation of innovative and improved versions of validated and accepted reference test methods that may provide advantages such as greater accuracy and efficiency. An appropriate set of high quality reference substances that can be readily obtained is critical for adequate performance standards. These reference substances should be selected from a robust database of commercially available candidate substances. Therefore, it is essential that stakeholders facilitate the collection and public availability of high quality reference substance data to support robust performance standards. Performance standards are now routinely

developed and incorporated in national and international guidelines and are consistent with international guidance on validation and regulatory acceptance as outlined in OECD Guidance Document 34 (OECD, 2005).

References

Basketter, D., Cockshott, A. et al. (2008). An evaluation of performance standards and non-radioactive endpoints for the local lymph node assay. *ATLA 36(2),* 243-257. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-45549100795&partnerID=40&rel=R8.0.0

Ehling, G., Hecht, M. et al. (2005). An European inter-laboratory validation of alternative endpoints of the murine local lymph node assay 2nd round. *Toxicology 212*, 69-79.

EPA, U. S. (1998). Health Effects Test Guidelines: OPPTS 870.1200 – Acute Dermal Toxicity. EPA 712-C-98-192. Washington, DC, U.S. Environmental Protection Agency. Available at: http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-1200.pdf

ICCVAM (2009). Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf (Accessed: September 17, 2009).

ICCVAM (2004). Recommended Performance Standards for In Vitro Test Methods for Skin Corrosion: Prepared by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 04-4510. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/dermal_docs/ps/ps044510.pdf (Accessed: September 17, 2009).

ICCVAM (2003). ICCVAM Guidelines for the Nomination and Submission of New, Revised and Alternative Test Methods NIH Publication Number 03-4508. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.pdf

ICCVAM (2002). Evaluation of EPISKIN™ , EpiDerm™ , and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: In Vitro Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication Number 02-4502. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/dermal_docs/cwgfinal02/cwgfinal.pdf (Accessed: September 17, 2009).

ICCVAM (1999). Corrositex®: An In Vitro Test Method for Assessing Dermal Corrosivity Potential of Chemcials - An Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 99-4495. Research Trinagle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/reports/corprrep.pdf (Accessed: September 17, 2009).

Idehara, K., Yamagishi, G. et al. (2008). Characterization and evaluation of a modified local lymph node assay using ATP content as a non-radio isotopic endpoint. Journal of Pharmacological and Toxicological Methods 58(1), 1-10. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-42649119427&partnerID=40&rel=R8.0.0 (Accessed September 17, 2009).

NIEHS (2008). Announcement of an Independent Scientific Peer Review Panel Meeting on the Murine Local Lymph Node Assay; Availability of Draft Background Review Documents; Request for Comments. Federal Register 73, 1360-1361. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_25553.pdf (Accessed September 17, 2009).

NIEHS (2007a). The Murine Local Lymph Node Assay: Request for Comments, Nominations of Scientific Experts, and Submission of Data. Federal Register 72, 27815-27817. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf. (Accessed September 17, 2009).

NIEHS (2007b). Draft Performance Standards for the Murine Local Lymph Node Assay: Request for Comments. Federal Register 72, 52130-52131. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_18011.pdf (Accessed September 17, 2009).

OECD (2009a). Draft Proposal for an Update to Test Guideline 430: In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER). Available at: http://www.oecd.org/dataoecd/14/2/43295353.pdf (Accessed September 17, 2009).

OECD (2009b). Draft Proposal for an Update to Test Guideline 431: In Vitro Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method. Available at: http://www.oecd.org/dataoecd/63/6/43302385.pdf (Accessed September 17, 2009).

OECD (2006). OECD Guidelines for the Testing of Chemicals. Test No. 435: In Vitro Membrane Barrier Test Method for Skin Corrosion. Adopted July 19, 2006. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s33/p1 (Accessed: September 17, 2009).

OECD (2005). OECD Series on Testing and Assessment. Number 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Available at: http://appli1.oecd.org/olis/2005doc.nsf/linkto/env-jm-mono%282005%2914 (Accessed: September 17, 2009).

OECD (2002). OECD Guidelines for the Testing of Chemicals. Test No. 429: Skin Sensitisation: Local Lymph Node Assay. Adopted April 24, 2002. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s29/p1 (Accessed: September 17, 2009).

Stokes, W. S. and Schechtman, L. M. (2007). *Validation and regulatory acceptance of new, revised, and alternative toxicological methods*. *Principles and Methods of Toxicology*. A. W. Hayes. Philadelphia, Pennsylvania: Taylor and Francis: 1103-1128.

Stokes, W. S., Schechtman, L. M. et al. (2006). The use of test method performance standards to streamline the validation process. *ALTEX 23, Suppl*., 342-345. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17492210

Takeyoshi, M., Yamasaki, K. et al. (2001). Development of non-radio isotopic endpoint of murine local lymph node assay based on 5-bromo-2?-deoxyuridine (BrdU) incorporation. *Toxicology Letters 119(3),* 203-208. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035826075&partnerID=40&rel=R6.0.0 (Accessed September 17, 2009).

Acknowledgements

The authors acknowledge the expert assistance of David Allen and Steven Morefield from ILS, Inc., the NICEATM Support Contractor, and Debbie McCarley, NICEATM/NIEHS, in the preparation of this manuscript. The content of this manuscript was supported by the Intramural Research Program of the National Institutes of Health.

Correspondence to

William S. Stokes, D.V.M.
Director, NICEATM
National Institute of Environmental Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park
NC 27709, USA
e-mail: stokes@niehs.nih.gov