



Center for
Scientific Review

Plans for the Evaluation of the ENQUIRE Program

Lia Fleming, MPH

Senior Public Health Analyst
Strategic Planning and Implementation Branch
Division of Planning, Analysis, and Information Management

Bruce Reed, PhD

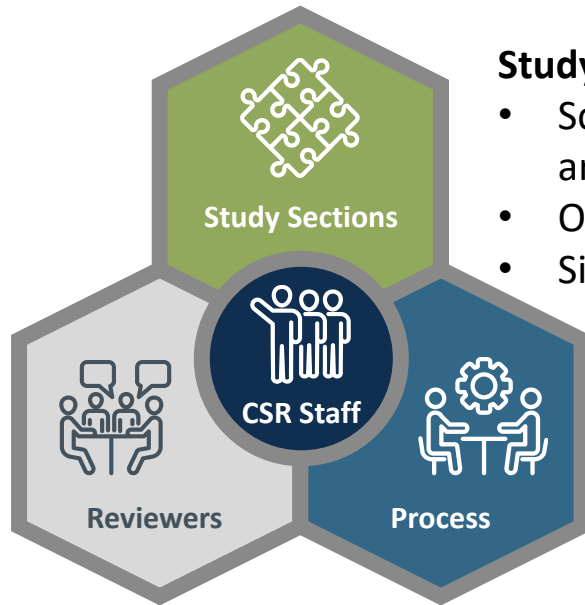
Deputy Director
Center for Scientific Review

September 23, 2024



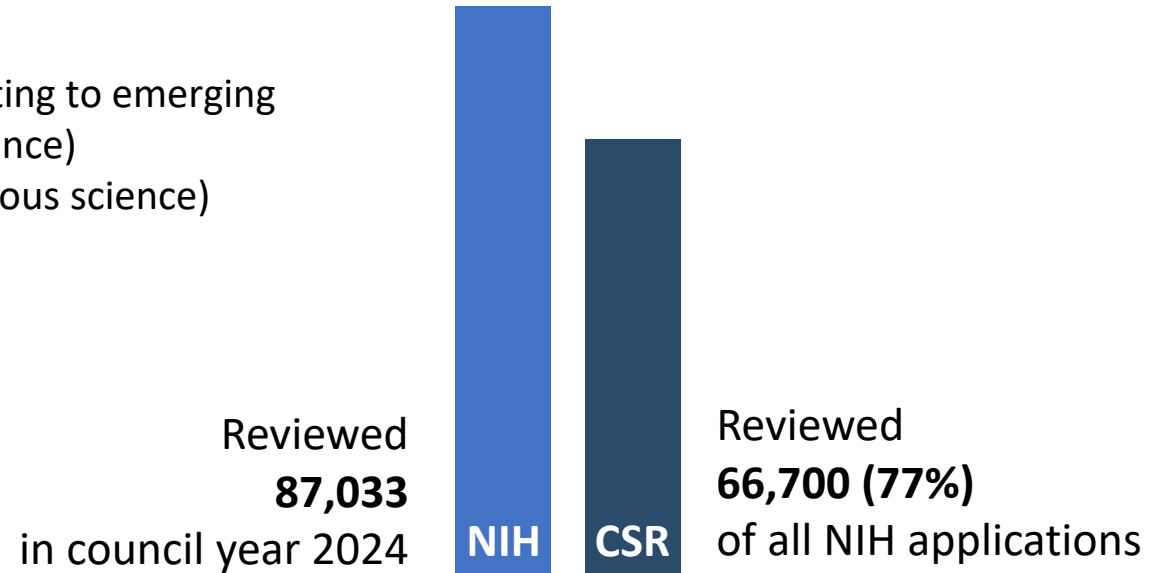
ENQUIRE
History and Background

Keeping study sections scientifically current across the breadth of NIH supported research is a fundamental challenge



Study Sections

- Scientific scope (relevance, adapting to emerging areas, not perpetuating stale science)
- Output (identification of meritorious science)
- Size appropriate for competition



- CSR's study sections should cover the entire spectrum of NIH research project grants
- Must incorporate emerging science; should not perpetuate outdated topics/approaches
- Appropriate size; manageable sets that provide the right level of competition

Previous approaches to keeping study sections scientifically current

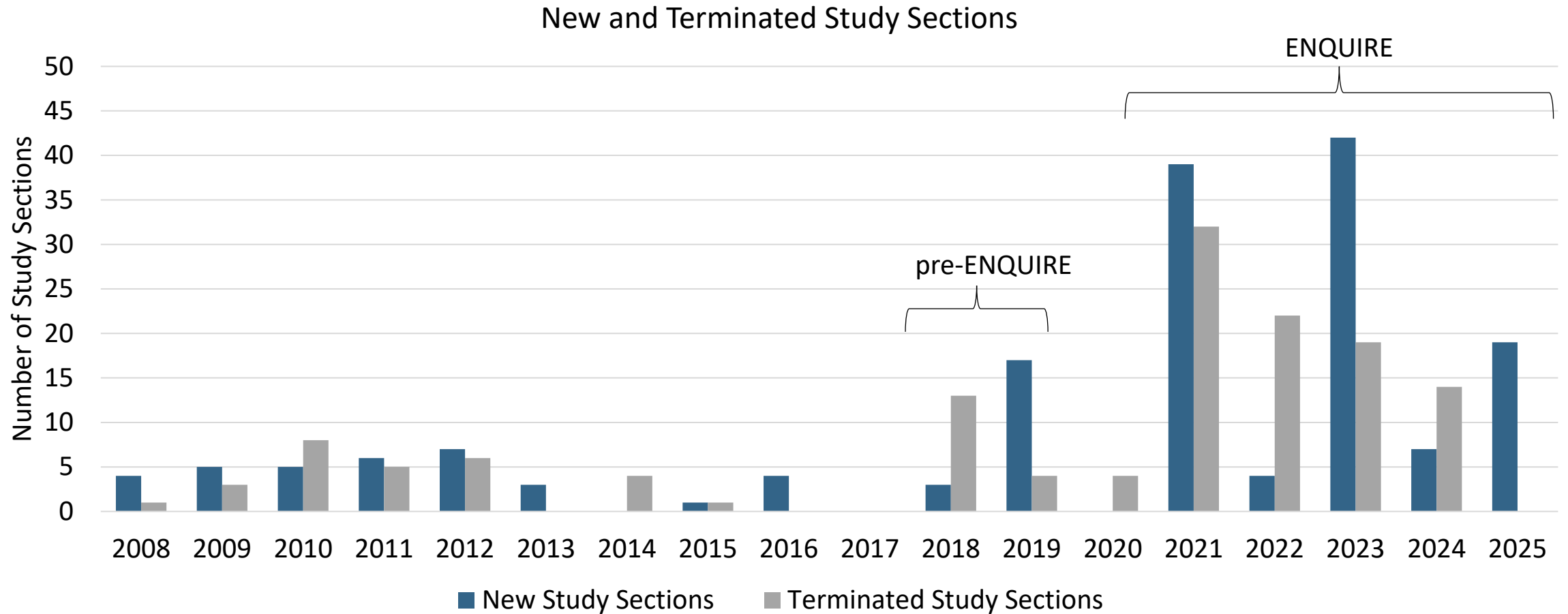
2003-2015

- Study sections evaluated in groups based on CSR management structure (review branch)
- Input from
 - CSR management/senior staff only (2004-2008)
 - Chairs, select reviewers (2008-2011)
 - Blue-ribbon external scientific working group (2011-2015)
- **Output was diffuse** – lots of comments conflating science and process (e.g. exit interviews, private discussions with SROs about management, should Chairs recruit reviewers, why did NIH take away the A2 and more)
- Only ***scientific*** changes to study sections were endorsements of proposals made by the CSR review branch chief during his/her presentation of the scientific scope

2016-2018 (Pre-ENQUIRE)

- Study sections evaluated in scientific groupings (**NOT** by review branch)
- Input from blue-ribbon external scientific working group provided with data on applications, bibliometrics, etc.
- **Output was significant scientific changes**, study sections restructured, eliminated, created
- **But no input from NIH program (ICs)**

Prior to ENQUIRE, there was very little change in CSR's study sections

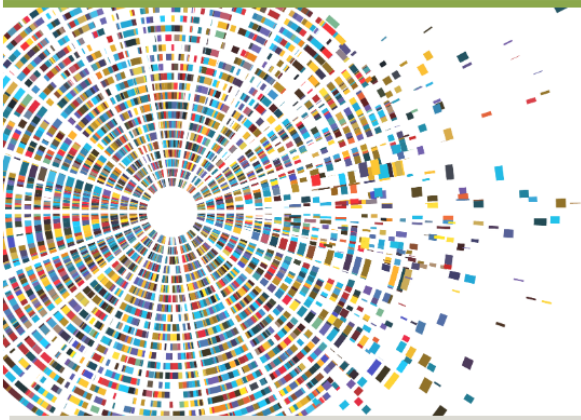


Since 2019, ENQUIRE – Evaluating Panel Quality in Review

- ENQUIRE integrates data and input from multiple stakeholders to evaluate the scientific scope and function of study sections
- Provides a systematic process for regularly evaluating all of CSR's 180+ study sections to adapt to changing scientific fields
- Study sections are clustered by field instead of CSR managerial structure
- Cycle 1 Goal; assess about 20% of study sections each year; review all study sections once every 5 years

Evaluating Panel Quality in Review (ENQUIRE)

[ENQUIRE Overview](#) | [Cluster Overview Process](#) | [Clusters Evaluated via ENQUIRE](#)



Science changes rapidly. Making sure that study sections change with the science is an ongoing challenge. CSR ENQUIRE integrates data and input from multiple stakeholders to determine whether changes in study section focus or scope are needed to facilitate the identification of high impact science, with special consideration of emerging science.


Words from Dr. Noni Byrnes, CSR Director
> [January 2020 NICHD Advisory Council meeting](#) [▶ WATCH](#)

ENQUIRE Overview

Science-based
Clusters of study sections are formed based on scientific topics (instead of CSR managerial units) for review via ENQUIRE

Systematic, data-driven, continuous process
~20% of CSR study sections are evaluated each year, every study section evaluated every 5 years

Stakeholder input and involvement



Multiple Possible Actions Follow

- **Change** in scientific guidelines
- **Merge** study sections
- **Create** new study sections
- **Redistribute** areas across study sections

Learn more: <https://public.csr.nih.gov/StudySections/CSREnquire>

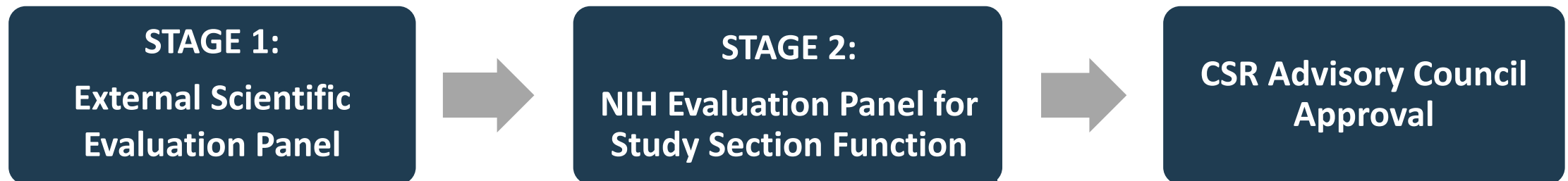
ENQUIRE uses a two-stage approach to evaluate study section science and function

Stage 1

- **External community input:** Selected members of scientific community with broad, relevant scientific expertise
- **Data provided:** e.g. output/publication data, early-stage investigator outcomes data, sample abstracts, & application load trends
- **Asked:** How well does the scope of the study sections align with the current state of the science? What is the optimal structuring of study sections to review the science?

Stage 2

- **NIH input:** Selected NIH program and review leadership with broad, relevant scientific interest
- **Data provided:** Stage 1 inputs plus score distributions, roster info/expertise, reports of meeting dynamics through study section site visits, program feedback surveys, recommendations from Stage 1, and more
- **Asked:** Comment on recommendations from external panel and consider any functional issues of existing study sections.



Wrapping up the 1st cycle of ENQUIRE

ENQUIRE SRG Progress (of CSR's 188 active study sections)



25 study sections will be deferred and prioritized to cycle 2 of ENQUIRE

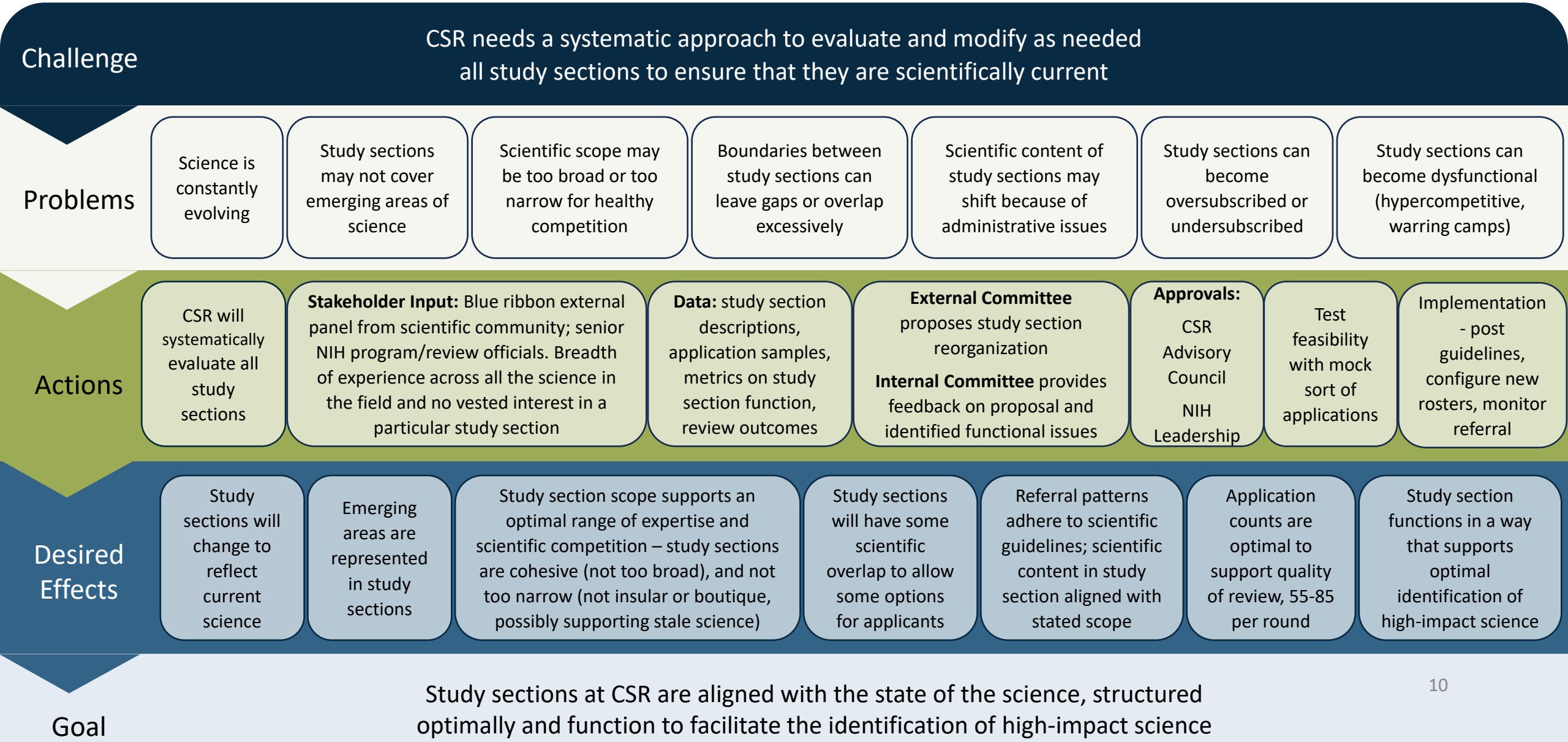
- 18 study sections – in the areas of HIV/AIDs, bioengineering, and imaging – were evaluated in ENQUIRE program precursors ~2018
- 2 recently-established study sections in cancer prevention and basic mechanisms in cancer health disparities – administrative considerations
- 5 genetics-focused study sections – active program changes

Complete streamlined evaluations of 6 remaining study sections 1 or 2 study sections at a time

- **Community input:** Recruit small group of external scientists with broad perspectives
- **Data:** Provide standard, external data package on these SRGs, including guidelines to overlapping clusters
- Provide set of standard, focused questions for panel to advise CSR on
- Intended as an interim measure until study sections can be reviewed fully in next cycle of ENQUIRE

Evaluating ENQUIRE

ENQUIRE Program Theory of Change



Why evaluate?

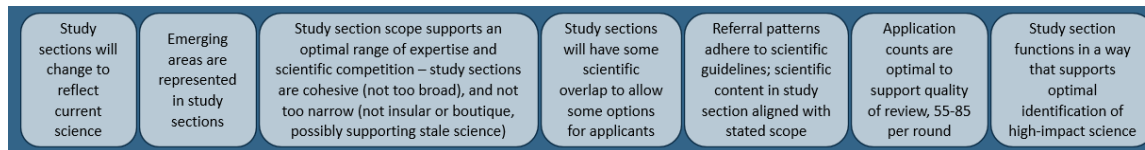
- 1 To grade past efforts holds us accountable and addresses stewardship responsibilities
- 2 To inform future efforts: As long as we have study sections, we must address the fundamental challenge of keeping study sections scientifically current. Each effort should be better than the last.

What should we evaluate?

Outcomes

- Important whether looking forwards or back
- Outcomes are generally what external parties are interested in

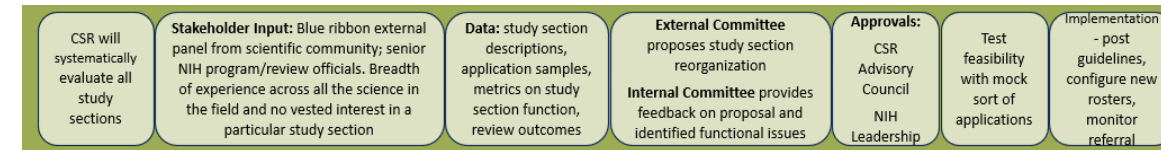
ENQUIRE “Desired Effects”



Processes

- Quality of implementation is relevant to quality of outcomes
- Critical for shaping future efforts—very important to CSR/NIH

ENQUIRE “Actions”



Evaluation Challenges in ENQUIRE

- **ENQUIRE is a complex, dynamic intervention implemented in a complex, adaptive system**
 - Cycle 1 will include 14 ENQUIRE reviews, done over 5 years. Iterative changes were to the program through lessons learned – processes were tweaked between reviews.
 - Each ENQUIRE “intervention” has multiple components (study section guidelines, associated referral changes, changes to rosters, changes in study section culture) that take effect at different times
 - Clusters were different at baseline, different in size, different issues in the study sections, different scientific contexts
 - Multiple significant changes to peer review between 2019 and now—new policies, new practices, new training
 - As science progresses applications change, reviewers change
 - Study sections are monitored and informally reviewed outside of ENQUIRE (changes made as needed)
- **Scale:** Over 180 study sections, 65k applications, 19k reviewers
- **No contemporaneous control group**
- **No “gold standard” outcome measure**
- **ENQUIRE is not a clinical trial**

ENQUIRE effects can be grouped into 3 sets:

Group 1: Three effects concern structural changes to study sections

Study sections will change to reflect current science

Emerging areas are represented in study sections

Application counts are optimal to support quality of review, 55-85 per round

Group 2: Two effects concern referral of applications to study sections

Study sections will have some scientific overlap to allow some options for applicants

Referral patterns adhere to scientific guidelines; scientific content in study section aligned with stated scope

Group 3: Two effects concern quality of review

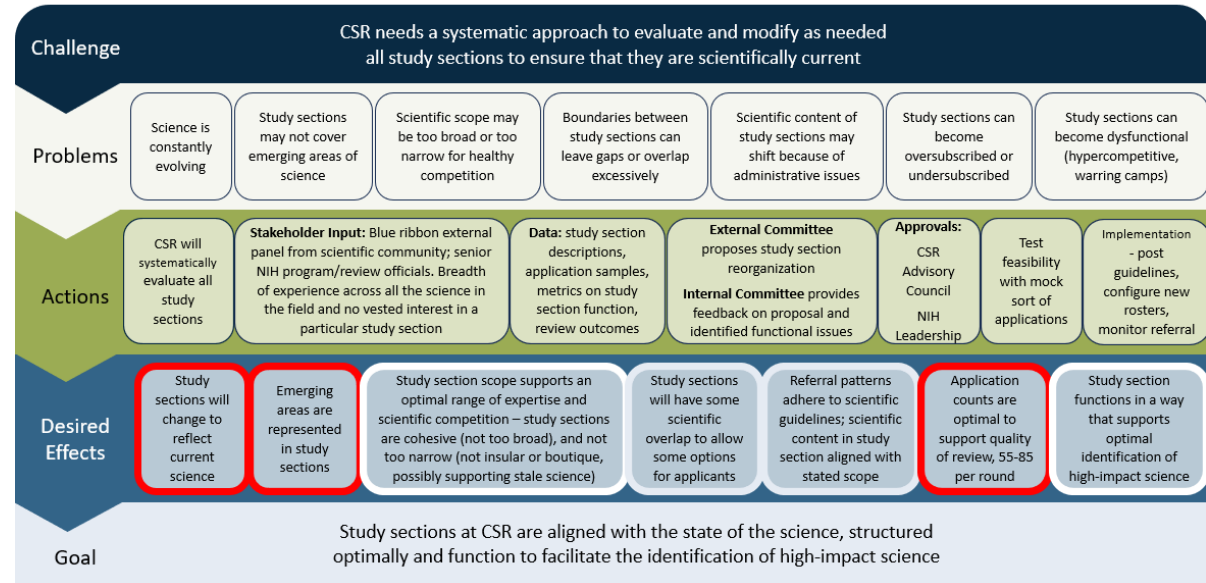
Study section scope supports an optimal range of expertise and scientific competition – study sections are cohesive (not too broad), and not too narrow (not insular or boutique, possibly supporting stale science)

Study section functions in a way that supports optimal identification of high-impact science

Group 1: Structural Changes

- Study sections changed to reflect current science
- Emerging science is covered well in CSR study sections
- Study sections review 55-85 applications per round

ENQUIRE Program Theory of Change



Group 1 (structural change):

Desired effect: Study sections changed to reflect current science

Counting structural changes:

- All study section guidelines changed post-ENQUIRE. Some were refreshed versions of the old study section; some were essentially new.
- Analysts, working with division directors classified each post-ENQUIRE study section as “substantially changed” or “updated”.
 - **“Substantial changed”** means the study section was essentially new, resulting from mergers and/or terminations, and/or moving sets of topics
 - **“Updated”** means the study section was refreshed--scientific guidelines were updated, topically emerging science was added

ENQUIRE resulted in substantial structural change to most study sections

11

clusters implemented

122

panels evaluated

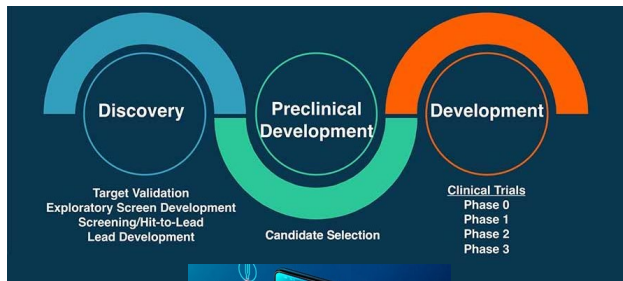
128

panels resulting

57%

of the resulting study sections were substantially changed

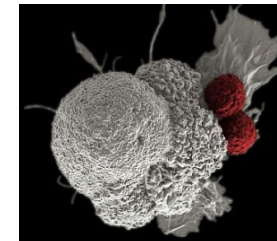
some examples....



Therapeutics: Late-stage preclinical drug discovery, biologics/drug delivery



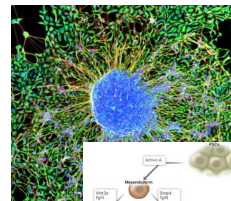
Social Determinants of Health



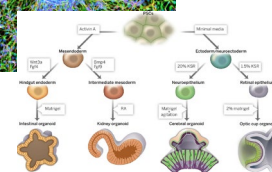
Cancer Immunotherapy



Mobile Health Technologies



Cell fate, stem cells and regeneration



Tissue formation and organogenesis

Group 1 (structural change):

Desired effect: Emerging science is covered well in CSR study sections

Challenging to assess, we plan two approaches:

1. Compile data on applications from Early Stage Investigators (ESI).
 - Baseline data exist. We will calculate values for post ENQUIRE study sections
 - “ESI” only crudely approximates “emerging science”
2. Survey applicants and reviewers on whether study sections guidelines encompass the new scientific topics of their field?
 - New survey items needed—no baseline data. Overall, best option.
3. More quantitative methods are difficult to scale and interpret
 - E.g. Some ENQUIRE reviews listed areas of emerging science; we could count their frequency in new study section guidelines. Using NLP models we could tag applications on those topics and analyze trends in application counts, and which study sections review them
 - » Tagging applications focusing on those topics using NLP/AI is possible, but requires considerable human effort; precise tagging is hard to achieve
 - » Informative? An increase in applications on a topic cannot be simply attributed ENQUIRE. An increase in applications may only reflect larger scientific trends.

Group 1 (structural change):

Desired effect: Study sections review 55-85 applications per round

Pertinent data are available. We track N of applications per study section each round.

Post-ENQUIRE changes in the number of meetings considered too large or too small										
Metrics	Cluster A		Cluster B		Cluster C		Cluster D		Cluster E	
	Pre-ENQUIRE	Post-ENQUIRE	Pre-ENQUIRE	Post-ENQUIRE	Pre-ENQUIRE	Post-ENQUIRE	Pre-ENQUIRE	Post-ENQUIRE	Pre-ENQUIRE	Post-ENQUIRE
#/% of Meetings >100	0/ 0.00%	1/ 3.57%	0/ 0.00%	0/ 0.00%	0/ 0.00%	0/ 0.00%	1/ 2.27%	0/ 0.00%	12/ 33.33%	5/ 11.36%
#/% of Meetings <50	8/ 23.53%	0/ 0.00%	26/ 40.63%	0/ 0.00%	4/ 11.11%	1/ 2.63%	16/ 36.36%	9/ 18.75%	0/ 0.00%	0/ 0.00%

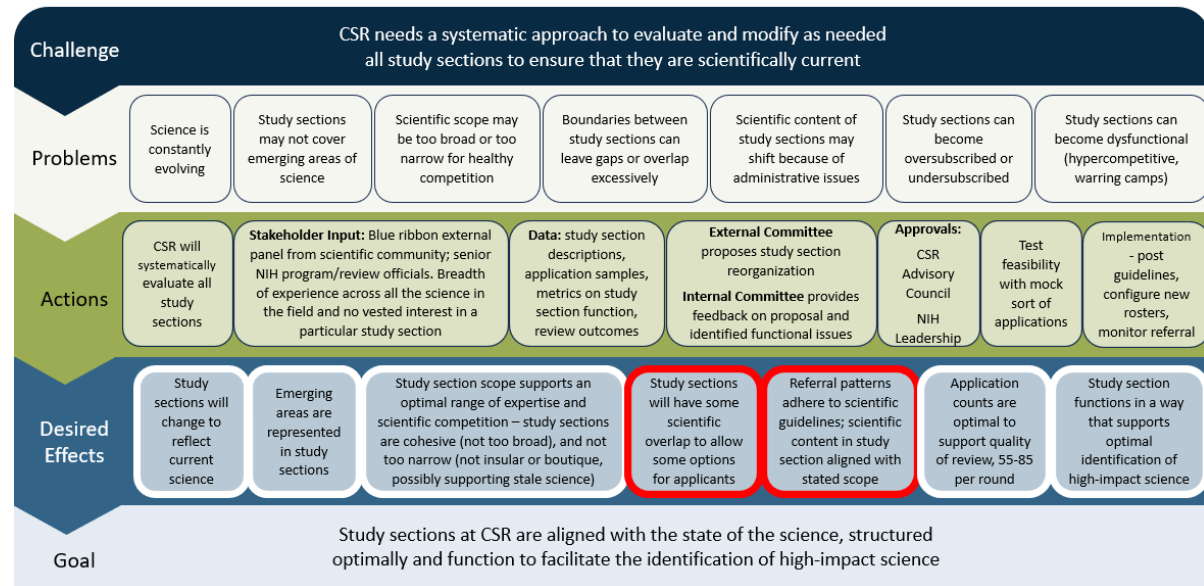
Caveats regarding changes in study section loads

- More study sections have application counts in the target range now than pre-ENQUIRE
- ENQUIRE likely contributed to this improvement but management practices have a major effect, e.g.
 - CSR will not charter study sections that are too small
 - When panels get too many applications in a round, overflow SEPs are created
 - If a chartered study section is chronically too large, CSR will divide it
- Maintaining target study section size is not simply managing workloads, it's keeping size appropriate for high quality review

Group 2: Application Referral Patterns

- Study sections will overlap scientifically, giving applicants options
- Referral patterns adhere to scientific guidelines

ENQUIRE Program Theory of Change



Group 2 (application referral patterns):

Desired effect: Study sections will overlap scientifically, giving applicants options

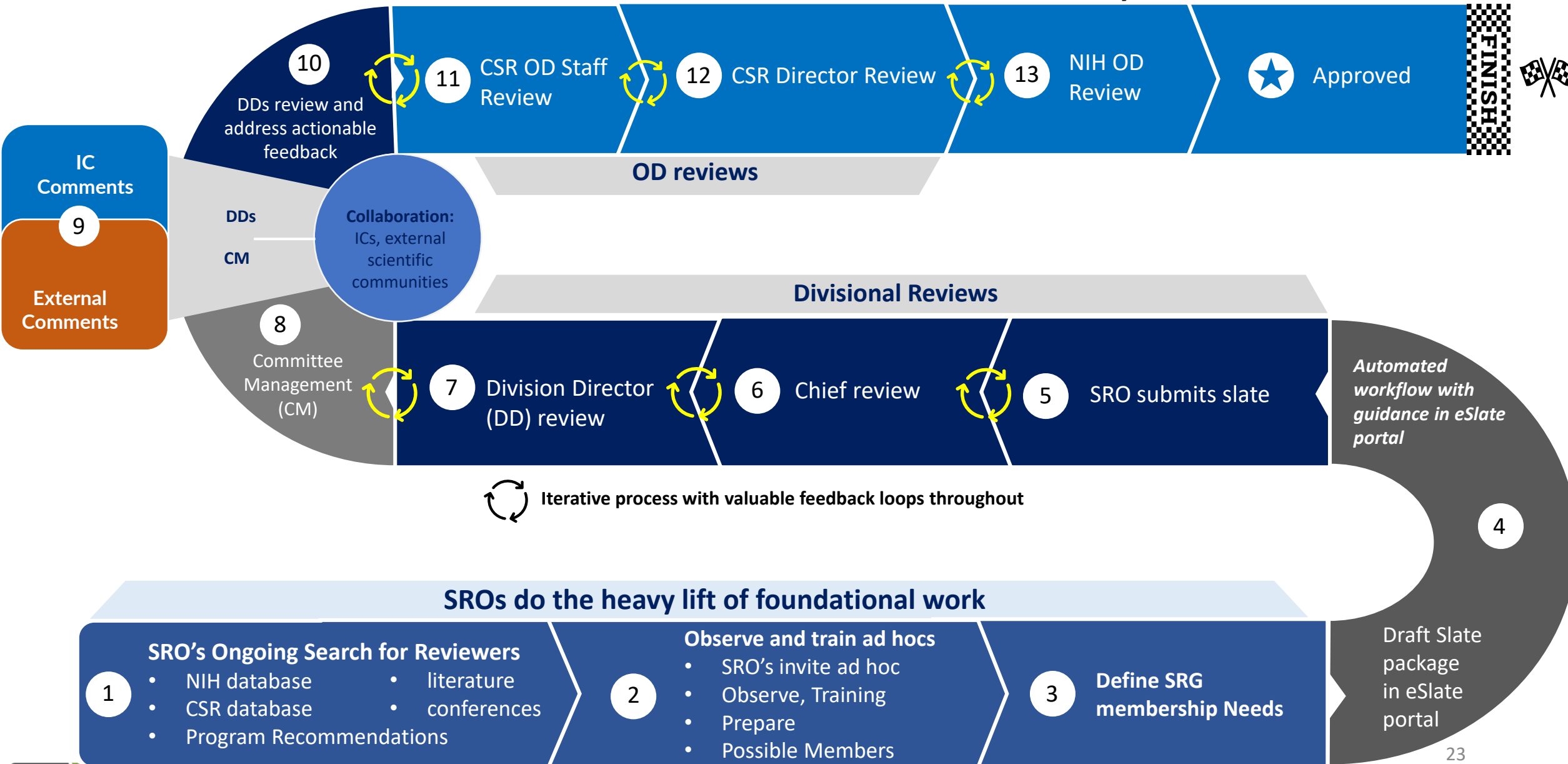
- Challenging to operationalize the desired degree of overlap, challenging to measure overlap informatively
 - No feasible measure in our administrative data
 - There are computational approaches to measuring concept similarity but interpretation is unclear
- Overlap is inevitable. Topics boundaries are fuzzy, the goals, models, methods, and different aims of an application may point to different study sections. Differentiating study sections is more of a challenge than is creating overlap.
- CSR pays attention to overlap. Every application needs to be assigned. There are many situations where overlap is desirable (COI, PI requests, locus of review commitments).
- Formal evaluation is challenging; CSR evaluates overall informally on an ongoing basis.

Group 2 (application referral patterns):

Desired effect: Referral patterns adhere to scientific guidelines

- Deviations from published guidelines are problematic.
 - Following new guidelines is necessary to cement change
 - Transparency, fairness to applicants
 - Quality of review. Guidelines were designed to promote healthy competition so deviations are undesirable.
- Measuring fidelity of referral at scale is very difficult.
 - Expert evaluation—e.g. *post-hoc* review of assignments by a DD or branch chiefs is very time consuming
 - AI? We have not seen a viable approach
- The nomination slates process provides a check—both internal and external experts examine fit of nominees to slate guidelines.

Reviewer Recruitment & Selection: The Slate Journey



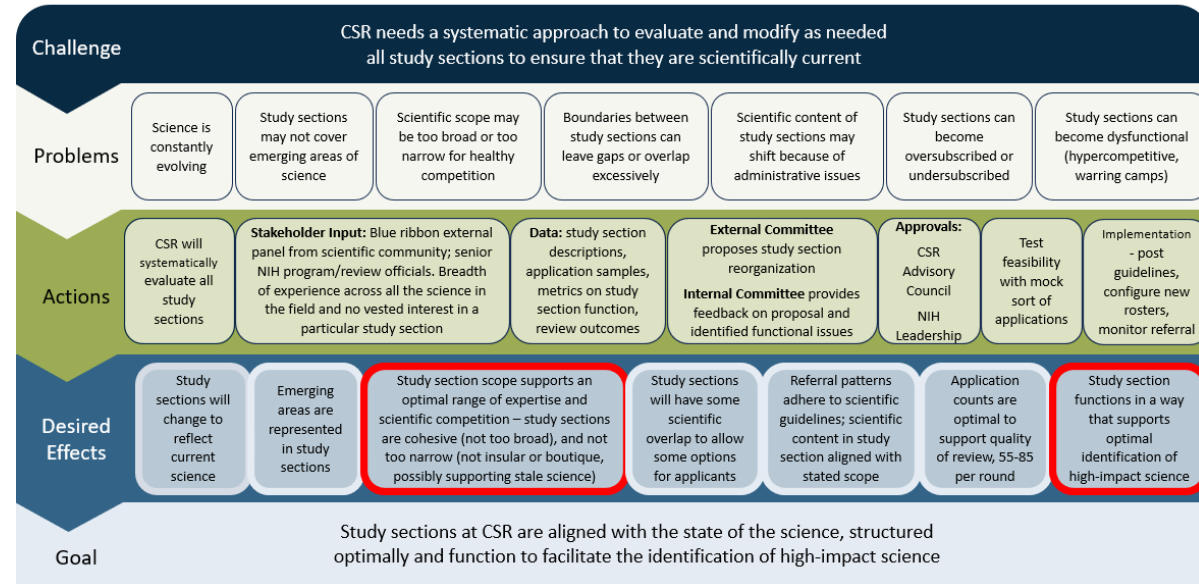
Group 2 (application referral patterns): Summary

- Formal evaluation of overlaps and fidelity of referral is difficult. We have not identified a feasible formal or quantitative approach.
- The slates nomination process includes external evaluation check on roster expertise and appointing the right expertise to the panel provides some assurance that the guidelines are being followed.
- Both of these outcomes relate to a core CSR process—the referral of applications to study sections. Both receive ongoing internal attention and informal evaluation from staff at multiple levels.

Group 3: Quality of Review

- Study sections should provide a healthy competitive context for review
- Study sections should identify the best science—the applications that will have the greatest scientific impact.

ENQUIRE Program Theory of Change



Group 3 (quality of review):

1. Study sections should provide a healthy competitive context for review.

- Not too broad— need to be cohesive, so reviewers can inform each other through discussion
- Not too narrow—insular, “entitlement fields”, possibly scientifically stale

2. Study sections should identify the best science—the applications that will have the greatest scientific impact

The question: “Do study section do a better job of identifying high impact science now than they did pre-ENQUIRE?”

- No objective, scalable, well-accepted measure of the quality of study section output

Group 3 (quality of review):

Possible Approaches:

1. Bibliographic measures (RCR) or bibliographic plus other indicators like patents, treatment, or clinical guidelines
 - Study section level RCRs were provided ENQUIRE panels
 - Data take years to emerge, interpretation is fraught
2. Surveys stakeholders regarding study section function
 - Reviewers—
 - a long-standing practice; we have historical data to reference
 - Applicants—
 - no pre-ENQUIRE data, but we propose to implement

Past surveys have polled reviewers on key concepts

1. Quality of review:

(How well did the panel prioritize applications according scientific merit?)

2. Expertise coverage:

(Was the proper scientific expertise present?)

3. Appropriateness of assignments

(Were assignments of applications to reviewers appropriate?)

4. Quality of discussions

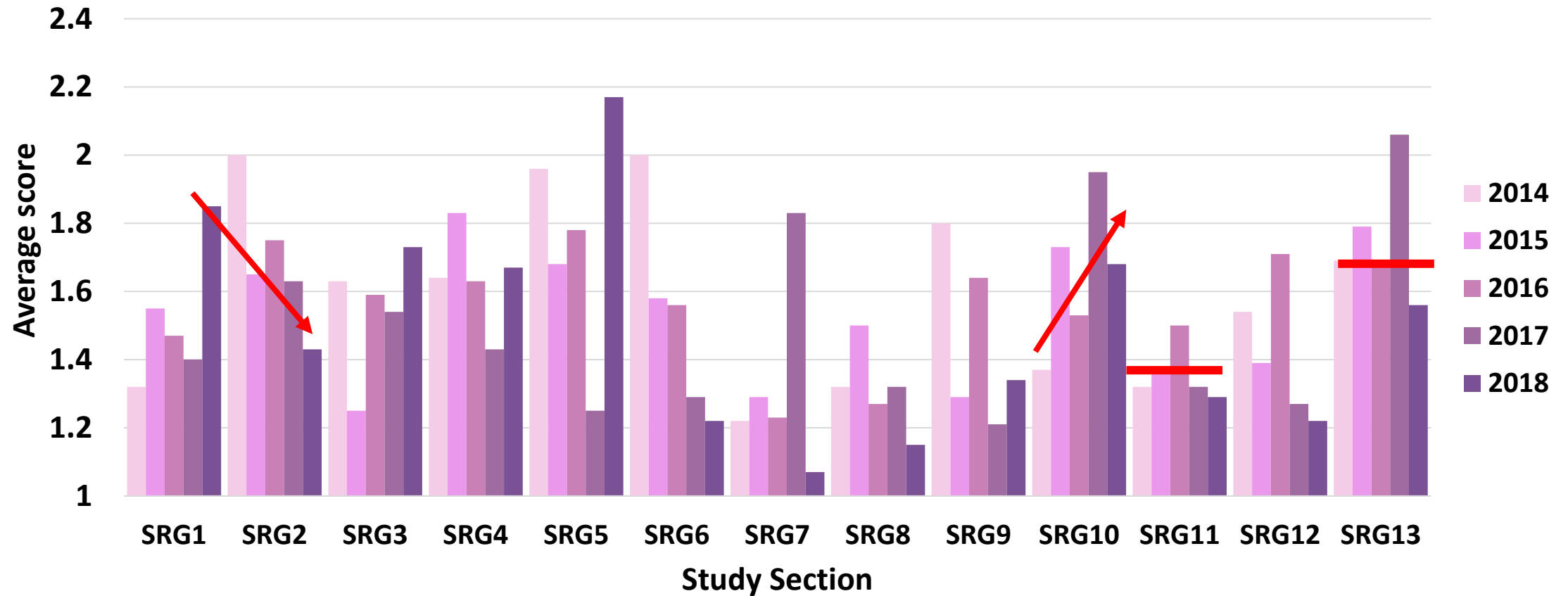
(Were discussions productive?)

5. Study section scope

(Is the scientific scope of the study section appropriate?)

Sample Data: 2 review branches polled 2014-2018

Q2 - Collective Expertise: The roster of reviewers was an appropriate assembly of scientific expertise for the set of applications in the meetings



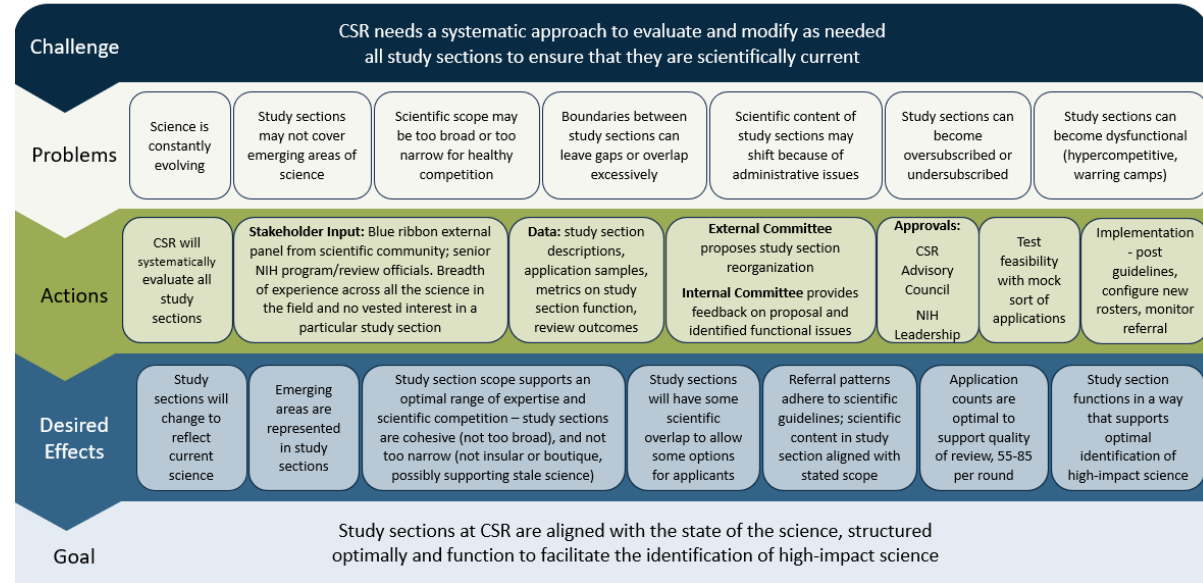
Caveats

- Happy reviewers does not equal quality review.
- Applicant responses are likely to be highly influenced by the outcome of their most recent grant submission.
- Change is not always welcome. Even necessary and desirable changes can be difficult in transition.

II. Evaluating ENQUIRE procedures

- Evaluating processes matters because quality of execution shapes quality of outcomes
- Evaluating past processes is critical for shaping future efforts

ENQUIRE Program Theory of Change



Approach to evaluating ENQUIRE procedures

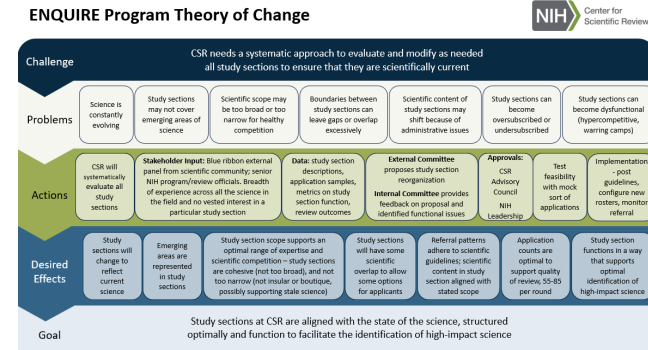
- CSR has informally evaluated each round of ENQUIRE. Those evaluations have had an impact: iterative improvements to ENQUIRE procedures, and the creation of resources to facilitate effective ENQUIRE reviews
- A formal, comprehensive process review would be resource-intensive.
- Instead, we plan two additional focused evaluations:
 1. Evaluate the data given ENQUIRE panels
 - Review requests for additional data we received from past panels. Do they point to additional standard data we should provide?
 - Survey ENQUIRE reviewers for feedback on data provided. What was useful, what was not, what additional data would they have liked?
 2. Evaluate whether NLP approaches can help with grouping study sections together for ENQUIRE review. Each ENQUIRE review evaluates a cluster of study sections—so defining those is fundamental.
 - Previously done by CSR senior staff—might NLP approaches be helpful?

An NLP approach to creating ENQUIRE clusters

- BioSimCSE Machine Learning model—dedicated tool for computationally characterizing science content-- was trained on 28 million publications in PubMed and their MeSH terms
 - Produces 768-dimension vectors that characterize each application
 - BioSimCSE Machine Learning model was applied to all RPGs reviewed in standing study sections over several rounds (49k applications), thus creating a high dimensional vector representing the scientific content of each application.
- The set of application vectors for each study section used to define a high dimensional vector characterizing the scientific content of every standing CSR study section.
- Study sections (vectors) were then grouped into 19 clusters according to vector similarity using a K-means approach. Resulting clusters submitted to CSR senior staff for review
- Many clusters looked fine as presented. Some had a stray member. One was too large and a couple were smaller than we like.
- Conclusion: starting with machine-generated clusters is a better approach.

Summary -1-

- CSR is wrapping up cycle 1 of ENQUIRE.
- ENQUIRE is a complex program, implemented on the background of significant changes to peer review and to science itself.
- A theory of change model guided our evaluation. TOC defines the component processes of ENQUIRE and connects the problems that follow from leaving study sections unchanged to the desired effects of the ENQUIRE program.
- **ENQUIRE processes** have been evaluated from the start. As we plan for ENQUIRE 2.0 we will obtain feedback from ENQUIRE participants and anticipate limited use of AI to facilitate the process.
- The **desired effects of ENQUIRE** form three bundles: effects related to Structural Change, to Referral, and to Quality of Review



Summary -2-

- Data show that ENQUIRE resulted in **extensive structural changes** to study sections across CSR. Additional evaluations:
 - monitoring frequency and outcomes of applications from ESIs
 - Surveys of reviewers and applicants on study section scope and coverage of emerging science
- Outcomes pertaining to **referral** do not lend themselves to evaluation metrics.
 - However, referral is a central business process of CSR, and is informally evaluated on an ongoing basis.
- **Quality of review** matters most but can be only incompletely evaluated.
 - Surveys are one indicator, with known flaws, which we propose to use.
 - They cannot substitute for thoughtful attention of SROs, Branch Chiefs, and CSR senior staff.

Questions for Advisory Council

- Thoughts on our framework and plan?
- Are we asking the right questions?
- Are there additional outcomes or evaluation approaches we should consider?
- Thoughts on how to communicate ENQUIRE evaluations?

END OF PRESENTATION