SGMA

Scientific Governance and Measurement Authority

POPULATION STANDARDS · EVIDENCE & IMPLEMENTATION

Comprehensive Case for SOGI Data Collection in Clinical Research

The consolidated scientific and operational case for collecting sexual orientation and gender identity data across clinical research.

Version 2.0 · 2026 · Active

Document	Comprehensive Case for SOGI Data Collection in Clinical Research
Version	2.0 (2026)
Status	Active
Publisher	SGMA (Scientific Governance and Measurement Authority)
Companion standards	SGMA-STD-001 (SOGI Protocol Implementation Standards) · SGMA-STD-002 (Population Classification Standard)
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
Recommended citation	SGMA. Comprehensive Case for SOGI Data Collection in Clinical Research. Version 2.0. SGMA; 2026.

Authors

Lauren B. Beach, JD, PhD · Amanda Beasley, PhD, MS · Craig Davenport · Terttu E. E. Haring, MD · Jackie Kent · Garo Kiledjian · Evan Ko · Kelly McKee · Meghan McKenzie, MA · Lisa Moneymaker · Madeline Ormand · Jirair Ratevosian · Donna Sattler · Kemi Williams, MBA · Benjamin Young, MD, PhD, FIDSA

Authors listed alphabetically by surname. Authorship was determined under SGMA-GOV-PP-001 (Publications & Presentations Authorship Standards), which applies ICMJE criteria.

Funding Disclosure

SGMA (Scientific Governance and Measurement Authority, pronounced “sigma”) is supported by foundation grants, individual donors, and institutional sponsors. Sponsor organizations do not direct, review prior to publication, or exercise editorial control over the scientific content of SGMA publications. The content and conclusions of this document are the independent determination of SGMA.

Executive Summary

Sexual orientation and gender identity (SOGI) data are not routinely collected in clinical research. This is more than an absent demographic field. It limits what trials can detect about safety and efficacy, what regulatory submissions can characterize, and what real-world practice can be informed by.

Demographic capture has long shaped clinical science. Age, sex, race, and ethnicity all gave researchers tools to detect differences that mattered for safety, dosing, and treatment response. SOGI sits in that same lineage, but with a measurement gap that has persisted across most therapeutic areas. Trials are still designed using protocol language, eligibility criteria, and case report forms built around an assumption of binary sex. The populations affected are not small. Recent surveys put sexual and gender minorities (SGM) identification in U.S. adults at 9% to 9.3%, with rates above 20% among adults under 30.[1–2] Internationally, identification reaches 17% among Generation Z across 26 countries, the highest of any generation, with 11% among millennials.[2–4]

Major pharmaceutical companies have established precedent. Bristol Myers Squibb began voluntary SOGI data collection in U.S. adult clinical trials in 2020. Genentech published peer-reviewed feasibility data on SOGI capture across clinical research and patient assistance program settings. GSK requires representation plans across all Phase 3 trials, with enrollment thresholds set against ethnicity, sex/gender, race, and age tied to disease epidemiology (SOGI is not yet part of those published criteria). Johnson & Johnson's LIBERTAS prostate cancer study used degendered, transgender-inclusive eligibility criteria. The Clinical Data Interchange Standards Consortium (CDISC) has published SOGI data standards.[3–7]

Across five disease areas where SGM populations face elevated or proportional disease burden, infectious diseases, mental health, hypertension, autoimmune disorders, and type 2 diabetes, treatment cost burden attributable to SGM populations is significant globally, in the multi-trillion-dollar range across these areas combined (see Methodology, Appendix A). The figure reflects treatment cost in populations whose data is largely absent from trial datasets generating treatment evidence, not market opportunity or revenue projection. Its scientific significance lies in the asymmetry: trials generating the evidence base for these therapeutic areas collect no systematic SOGI data, foreclosing safety subgroup analysis, hormone therapy drug-drug interaction assessment, and regulatory submission completeness.

Regulatory framing is shifting in real time. The FDA Diversity Action Plan draft guidance, originally issued April 2022 and updated in June 2024 under the statutory authority of the Food and Drug Omnibus Reform Act (FDORA, 2022), was withdrawn in January 2025 and restored to the FDA website in February 2025 by court order. The restored page carries an administrative disclaimer indicating the page is restored per court order and that the current administration rejects content related to gender identity. The FDORA statutory deadline for finalizing the guidance, June 26, 2025, was missed. The HHS Secretary committed during January 2025 confirmation testimony to finalize the guidance, though no timeline has been published. The underlying FDORA statutory requirements remain in force.[8–10] Outside the U.S., the Declaration of Helsinki, ICH E6 (Good Clinical Practice), the European Medicines Agency, and the U.K. Health Research Authority continue to require accurate study population characterization. Pharmaceutical companies operating across jurisdictions face an alignment problem that does not wait on a single regulator's posture.

This document gives clinical research and healthcare professionals at pharmaceutical companies the consolidated case for SOGI data collection. Implementation specifications are set out in SGMA-STD-001 (SOGI

Protocol Implementation Standards) and the controlled vocabulary in SGMA-STD-002 (Population Classification Standard).

Note on the Evidence Base

SGMA publishes this document drawing on peer-reviewed literature, regulatory guidance, and industry-documented practice. Where evidence is limited, that limitation reflects the underlying data gap this document addresses. Building the rigorous SOGI evidence base across therapeutic areas is an active priority of SGMA's Standards Governance and Implementation Committee. This document will be updated as the evidence matures, with a revised economic methodology under development with academic partners for inclusion in a future revision.

Key Terms and Concepts

Terminology in this document conforms to the controlled vocabulary in SGMA-STD-002 (Population Classification Standard). The cisgender heterosexual population is the formal full term for the comparator group; non-SGM population is the approved shorthand after first definition.

SOGI	Sexual Orientation and Gender Identity. The standard measurement framework used across clinical research protocols, regulatory guidance (FDA, EMA, ICH), global health frameworks (WHO), and healthcare data standards (CDISC). Used here for data and measurement references.
SGM	Sexual and gender minorities. The population-level term, used after full definition on first reference. Per SGMA-STD-002, SGM is defined analytically from the population-complete dataset, not pre-selected at enrollment.
Cisgender heterosexual population	The subset of study participants who self-report as both cisgender (gender identity corresponds to sex assigned at birth) and heterosexual, derived analytically from the population-complete dataset. The primary comparator group in SOGI-stratified analyses. Approved shorthand after first definition: non-SGM population.
Sex assigned at birth (SAAB)	Sex recorded at birth based on observable physical characteristics. A distinct variable from gender identity. Captured separately in CDISC SOGI codelists.
Gender identity	An individual's internal sense of gender, which may or may not correspond to sex assigned at birth. Self-reported.
Population-complete ascertainment	A data collection requirement in which SOGI variables are systematically captured from all enrolled participants regardless of demographic identity or risk classification. Per SGMA-STD-002, this is the prerequisite for any study claiming SOGI-informed analysis.
Social drivers of health (SDoH)	Economic, social, environmental, and geographic factors that influence health outcomes. SGMA uses 'drivers' rather than 'determinants' to reflect the modifiable nature of these factors.
Patient-centered access	Care that respects individual patient preferences without quality variation based on personal characteristics or social factors.

Precision medicine	Treatment approaches that account for individual characteristics, genetics, environment, lifestyle, in selecting therapies.

I. The Importance of Epidemiology in Medicine

Demographic data has driven clinical science for decades. Race and ethnicity data have helped researchers understand how disease is shaped by both genetics and the social conditions in which people live. Age and sex have informed safety, dosing, and outcomes for far longer. When demographic data is incomplete, science loses its ability to characterize variation across populations, and clinical care follows from what science can characterize.

SOGI does not stand in isolation. It interacts with race, socioeconomic status, geography, and other variables to shape clinical outcomes. Research has documented that intersecting identities can compound effects on recruitment, retention, and trial access.[11] The interaction between SOGI and socioeconomic factors influences medication adherence and follow-up care. Capturing these data lets researchers identify patterns that would otherwise remain invisible. Failing to capture them does not eliminate the underlying variation, it only removes researchers' ability to see it.

Three examples illustrate what demographic data has made possible:

- **Diabetes and Alzheimer's:** The Centers for Disease Control and Prevention (CDC) research confirms that social drivers of health directly influence prevalence and outcomes for both. Genetics matter, but disparities in healthcare access and lifestyle disproportionately shape disease progression and treatment effectiveness.[12–13]
- **Asthma:** American Lung Association data documents significantly higher asthma rates among Black Americans, particularly Black men, driven by environmental exposure, air quality, and reduced healthcare access. Housing conditions and historical inequities compound severity.[14–15]
- **Uterine fibroids:** the American College of Obstetricians and Gynecologists (ACOG) and the Journal of the American Medical Association (JAMA) Network Open data show fibroids are more prevalent and more severe in Black women, with Black, Hispanic, and Asian patients diagnosed at younger ages and facing greater treatment access barriers.[16–17]

Demographic data has done this work for race, ethnicity, age, and sex. SOGI data, currently absent from most clinical datasets, represents a critical blind spot in precision medicine for the populations it would otherwise serve.[18–21] Globally, the affected population is large and growing, with 17% of Generation Z and 11% of millennials across 26 countries identifying as a sexual or gender minority, numbering in the hundreds of millions of adults worldwide whose treatment evidence is generated without this data.[24]

SGM populations experience disproportionately high prevalence across several disease areas. Hypertension and infectious diseases occur at materially higher rates compared to non-SGM populations, driven by social drivers, stigma, and limited access to culturally competent care. Figure 1 illustrates the global distribution of disease burden across these high-prevalence disease areas.[22–24]

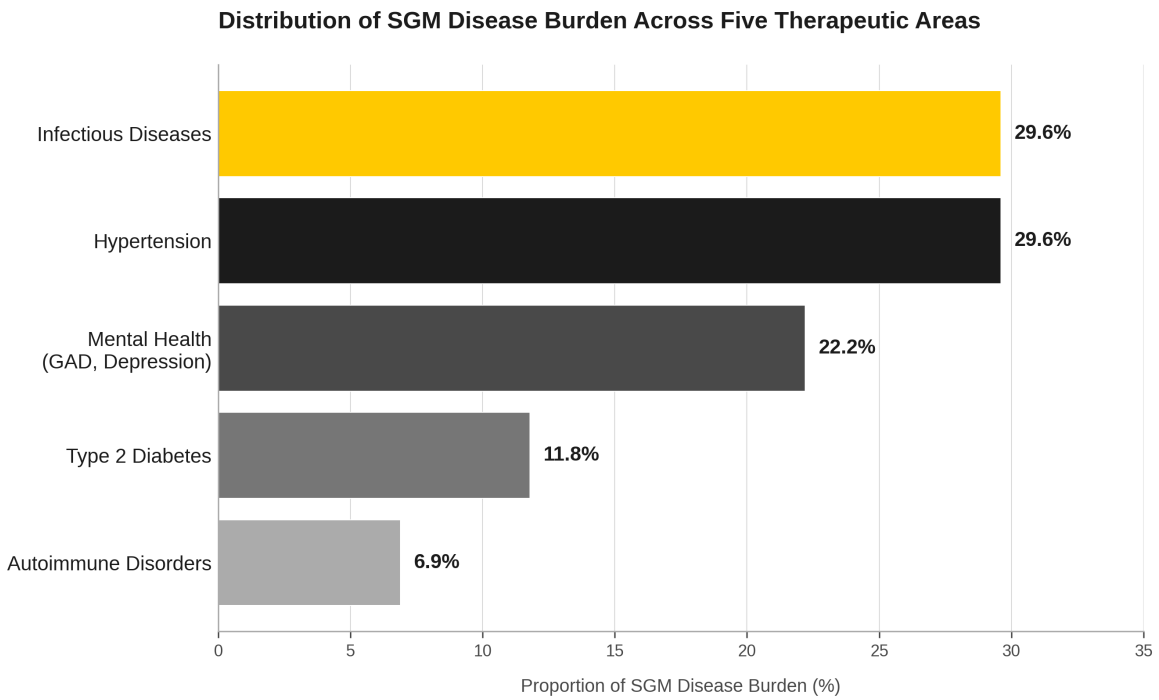

Figure 1. Distribution of SGM disease burden across five therapeutic areas. See Appendix A for methodology.

Pharmacological Relevance

Pharmacogenomic research has demonstrated that genetic variation across ancestries, alongside social drivers of health, contributes to differences in drug metabolism, efficacy, and safety. These differences shape regulatory approvals, market success, and commercial outcomes.[25–26] The NIH and pharmaceutical R&D programs have documented how genetic variation influences absorption, distribution, metabolism, and excretion, and ultimately effectiveness and adverse-event risk.[27–29]

While pharmacogenomic variation is primarily driven by ancestry, SOGI-related variables influence drug response through different but equally important mechanisms. Three are worth noting:

- **Hormone therapy as a concomitant medication.** Patients receiving exogenous hormones, for gender-affirming care, contraception, or menopause management, have well-documented PK/PD interactions with investigational products. When hormone therapy is not captured systematically as a concomitant medication, drug-drug interactions go undetected and adverse events are miscoded.
- **Physiologic accuracy.** Eligibility logic, safety monitoring, and subgroup analyses that assume binary sex variables can produce miscoded data when applied to populations whose sex assigned at birth and current physiologic status differ. Capturing SOGI alongside relevant physiologic variables resolves this.
- **Differential prevalence.** SGM populations experience materially **higher disease** prevalence in specific areas. Without SOGI data, subgroup analyses to quantify treatment response in these populations cannot be performed, because the data was never collected.

Without SOGI data, pharmaceutical companies risk developing therapies that overlook key patient populations and miss safety or efficacy patterns relevant to regulatory standards. Just as demographic data has reshaped understanding of racial and ethnic differences in disease and treatment response, SOGI data needs to be

included in clinical research routinely to produce the next generation of treatments that are both effective and safe across the populations using them.

PROVOCATIVE QUESTIONS FOR RESEARCH CONSIDERATION

- Are there measurable differences in prostate cancer incidence and treatment response between gay and non-gay men, and what factors explain those differences?
- Do lesbian women experience differential rates of breast cancer relative to heterosexual women?
- Are SGM individuals more likely to experience certain chronic illnesses due to stigma and healthcare access barriers?
- How do mental health clinical presentations differ in SGM populations relative to non-SGM populations?
- Are there clinically significant differences in medication efficacy or side effects in patients receiving exogenous hormone therapy?
- How do social drivers, homelessness, lack of insurance, interact with SOGI status to affect chronic disease outcomes?

SCHOLARLY SUPPORT

Comprehensive demographic data, including SOGI, has supported breakthroughs in drug development and clinical care, including in HIV.[18–20]

- **Goldhammer et al.** Research in the Journal of the American Medical Informatics Association establishes that SOGI data in electronic health records improves care and helps identify disparities affecting LGBTQIA+ populations.[20]
- **Jackson et al.** BMC Public Health research documents how stigma and discrimination contribute to distinct health challenges for sexual minorities, underscoring the case for systematic data collection.[30]
- **Shover et al.** American Journal of Public Health analysis demonstrates that absent SOGI data limits efforts to address disparities in HIV incidence among sexual minority populations.[31]

II. The Impact of SOGI Data Collection on the Healthcare Ecosystem

A. Strengthening Evidence-Based Medicine

Precision medicine works, and is only as effective as the data that reflects the patients receiving treatment. Capturing SOGI data expands what clinical research can see. With it, pharmaceutical companies and investigators can:

- Make a sizable population visible in datasets where it is currently invisible.
- Detect differences in treatment response and outcome in SGM populations.
- Ground precision medicine in the actual clinical considerations relevant to SGM patients, hormone therapy interactions among them.

- Expand therapeutic areas where SGM populations face higher prevalence: cardiometabolic, autoimmune, oncologic, mental health.
- Train machine learning and AI models on data that includes the populations those models will be applied to.

SCHOLARLY SUPPORT

- Schabath and colleagues found that institutional leadership and policy mandates significantly increased SOGI data collection in oncology, demonstrating both feasibility and clinical importance.[32]
- Karen Parker (NIH) has emphasized that systematic inclusion of SGM individuals improves generalizability (how well findings apply to populations beyond the trial sample) of clinical findings.
- Alpert and colleagues describe the SOGI-inclusive training and protocol design that supports equitable access to cancer trial participation.[33]
- Sorensen et al. point to a more practical barrier: inconsistent SOGI capture across clinical and registry systems, which constrains the population-based research that could otherwise narrow these gaps.[34]
- Trial recruitment that succeeds in SGM communities, where trust is earned by inclusion rather than assumed.

B. Strengthening Public Health Outcomes Globally

Beyond clinical trials, SOGI-inclusive data shapes public health practice across several connected fronts:

- Public health interventions that account for populations otherwise not represented in epidemiologic data.
- Therapies developed against SGM-relevant clinical considerations rather than against assumptions of demographic uniformity.
- Prescribing clinicians who have the data to make confident decisions across diverse patient populations rather than extrapolating from datasets that don't include those patients.

SCHOLARLY SUPPORT

- Pratt-Chapman's research demonstrates that normalizing SOGI data collection enables more precise interventions and builds an evidence base for future clinical advancement.[35]
- Scout's work on transgender health documents the role of SOGI data in improving care pathways.[36]
- Tran, Rosendale, and Lunn (JAMA Cardiology) outline the case for standardized SOGI data collection in cardiology research.[19]

C. Regulatory Continuity Across Jurisdictions

Regulatory guidance on demographic representation in clinical research has shifted in real time over the past two years. The FDA Diversity Action Plan draft guidance, originally published in June 2024 under the statutory authority of the Food and Drug Omnibus Reform Act (FDORA, 2022), was removed from the FDA website in January 2025 following Executive Order 14168. The guidance was restored on February 11, 2025 by court order. The restored page carries an administrative disclaimer indicating it is restored per court order and that the current administration rejects content related to gender identity. The FDORA statutory deadline for finalizing the draft, June 26, 2025, was missed. The HHS Secretary committed during January 2025 confirmation testimony to finalize the guidance, though no timeline has been published as of this writing.[8–9]

FDORA itself is law and remains in force. Regardless of administrative guidance volatility, the statute continues to authorize the Diversity Action Plan submission requirement once the draft guidance is finalized. The DAP framework addresses age, sex, race, and ethnicity; it does not require SOGI data specifically, so any pharmaceutical company collecting SOGI today does so voluntarily. Williams Institute analysis (February 2026) documents the broader pattern of federal data collection changes, including approximately 360 federal data collections that removed SOGI measures during 2025.[37]

Outside the U.S., the picture is less settled, not more. There is no instrument comparable to FDORA, and several jurisdictions treat SOGI as special-category data under data protection law (notably GDPR Article 9), which constrains collection rather than encouraging it. Under GDPR Article 9, processing sexual orientation data is prohibited by default unless an Article 9(2) exception applies, with the scientific research exception under Article 89 being the most operational route. This is precisely why a documented, auditable population standard matters: it gives pharmaceutical companies the framework to collect SOGI within these constraints rather than navigating them ad hoc. Canada's Sex- and Gender-Based Analysis Plus (SGBA Plus) framework, applied by federal funders including the Canadian Institutes of Health Research, requires integration of sex and gender considerations into research design; like other ex-US instruments, it stops short of mandating SOGI data collection specifically. The Declaration of Helsinki and ICH E6 (Good Clinical Practice) establish the requirement for accurate study population characterization across global research. The European Medicines Agency, the U.K. Health Research Authority, and Medicines and Healthcare products Regulatory Agency continue to apply guidance on inclusion and diversity in clinical studies submitted under their respective pathways.[38] For pharmaceutical companies operating across jurisdictions, the case for SOGI data collection rests on this multi-jurisdiction alignment more than on the posture of any single regulator at any single moment.

The argument for continuity rests on several factors:

- **Statutory authority persists.** FDORA and equivalent ex-US statutes remain in force regardless of administrative guidance volatility.
- **Industry adoption is documented.** Bristol Myers Squibb, Genentech, GSK, and Johnson & Johnson have established precedent for SOGI-inclusive trial design itself (eligibility criteria, case report forms, and protocol logic), not just trial materials as scientific practice, independent of regulatory cycles.
- **Global competitiveness.** Pharmaceutical companies with population-complete data infrastructure are positioned for global submissions where alignment with multiple jurisdictions matters.
- **Scientific rigor is non-partisan.** The case for accurate population characterization rests on study generalizability and pharmacovigilance, not on any particular policy posture.

D. Operational Implementation

Implementing SOGI data collection takes planning across several connected fronts. None alone is enough; together they move data capture from idea to infrastructure.

- **Participant trust.** Self-reported, voluntary collection works when the protocol is clear about how data will be used, secured, and protected. Communicate this once at consent and again wherever the questions appear.
- **Site staff readiness.** Clinical research coordinators administer SOGI questions every day they enroll. Sensitivity, consistency, and simple training resolve most operational friction. SGMA publishes guidance for pharmaceutical company and site teams.

- **Systems configuration.** CTMS (clinical trial management systems), EDC (electronic data capture) platforms, and specialized vendor systems accept SOGI variables when configured against the CDISC SOGI codelists. The CDASH (Clinical Data Acquisition Standards Harmonization)/SDTM (Study Data Tabulation Model) mapping is established; the integration into existing data systems is a routine configuration task, not a research problem.[6]
- **Multi-jurisdiction alignment.** Pharmaceutical companies running global programs coordinate SOGI capture against regional consent and data protection rules. Coordination with WHO and ICH frameworks supports cross-border consistency.
- **Implementation partnerships.** SGMA, CDISC, and other standards bodies publish frameworks that reduce the per-protocol cost of SOGI capture. Reuse the work that's already done.

E. Planning for Technology Innovation

Clinical data collected today becomes the historical dataset on which tomorrow's models are trained. Standardizing SOGI data collection now expands SGM representation in the data that will train clinical AI models, across recruitment optimization, synthetic control arms, and outcomes analysis. Without SGM data in training sets, these models cannot perform reliably for SGM populations, and outputs cannot be validated against differential performance.

III. Industry Milestones and Opportunities for SOGI Data

A. Industry Leadership

Several pharmaceutical companies have demonstrated that systematic SOGI data collection is operationally feasible and scientifically valuable. Their work establishes that population-complete trial design is not an aspirational concept but a documented practice.

- **Bristol Myers Squibb (BMS).** BMS established the first publicly documented voluntary SOGI data collection program in U.S. adult clinical trials in 2020. The implementation methodology, published the same year, remains a foundational reference in the field.[39]
- **Genentech.** The Phase III EMPACTA study (Genentech, tocilizumab in COVID-19 pneumonia) enrolled 389 participants from sites deliberately selected at hospitals serving high-risk and minority communities across the United States, South Africa, Kenya, Brazil, Mexico, and Peru. Approximately 85% of participants were from minority racial and ethnic groups, with majority Hispanic and significant Native American and Black representation, demonstrating that population-responsive trial design could support both scientific completeness and operational efficiency. Genentech subsequently published peer-reviewed feasibility data on SOGI capture across clinical research and patient assistance program settings.[40–41]
- **GSK.** Beginning in 2022, GSK implemented representation plans for every Phase 3 trial before enrollment, with thresholds set against ethnicity, sex/gender, race, and age tied to the disease epidemiology under study. GSK reported 100% of Phase 3 interventional trials initiated in 2023 having proactive diversity plans, and its 2025 inclusion reporting describes ongoing monitoring against those thresholds. SOGI is not yet part of GSK's published representation criteria, which is why a population standard such as SGMA-STD-001 is needed to extend this approach to sexual and gender minorities.[42]

- **Johnson & Johnson (J&J).** The LIBERTAS prostate cancer study applied degendered eligibility criteria and gender-neutral language, opening participation to transgender, nonbinary, and gender-diverse individuals.[43]

B. Data Standards Organizations: CDISC

The Clinical Data Interchange Standards Consortium (CDISC) has published SOGI data standards supporting consistent representation across regulatory submissions. The CDISC SOGI codelists capture sex assigned at birth, gender identity, and sexual orientation as distinct variables, providing the data architecture baseline for pharmaceutical company implementation. SGMA-STD-001 extends the CDISC infrastructure into protocol design, eligibility logic, pharmacovigilance integration, and regulatory submission templates.[6]

C. American Medical Association

The American Medical Association (AMA) supports inclusion of SOGI questions in electronic health records and has published guidance for clinical staff training in SOGI data collection. The AMA has also extended voluntary SOGI data capture to its own physician membership.[44–45]

D. Regulatory and Policy Framework

Multiple regulatory and policy bodies have established guidance and recommendations supporting SOGI data collection. The current landscape:

- **FDA.** The FDA Diversity Action Plan draft guidance published in June 2024; current status as described in Section II.[8–9] FDA guidance on sex and gender differences in clinical evaluation continues to support capturing sex-based and, where relevant, gender-based differences in safety and efficacy outcomes.[46]
- **National Academies of Sciences, Engineering, and Medicine (NASEM).** NASEM consensus reports recommend systematic SOGI data collection as a methodological standard for federal health research.[47]
- **International standards.** The Declaration of Helsinki, ICH E6 (Good Clinical Practice), the EMA, and the U.K. HRA/MHRA frameworks set expectations for accurate study population characterization in clinical research submitted under their respective pathways. None require SOGI data collection specifically, and in several jurisdictions SOGI is treated as a special category under data protection law.[38]

E. Contract Research Organizations (CROs): Implementation Partners

Leading CROs, including Syneos Health, Parexel, and PPD, have invested in clinical competency preparation and infrastructure capabilities supporting demographic data collection. ICON has operational capability for demographic data collection consistent with regulatory expectations. CROs are well-positioned to support pharmaceutical companies implementing SGMA-STD-001 across global trial programs.

CRO partnerships in support of SOGI data collection bring three operational benefits:

- **Operational consistency.** CROs can embed SOGI collection into protocol design uniformly across global sites, reducing data integrity errors.
- **Competitive differentiation.** CROs that adopt population-complete trial design position themselves as partners for pharmaceutical companies with global submission programs.
- **Regulatory support.** CROs equipped with SOGI-inclusive practices may help pharmaceutical companies meet evolving regulatory recommendations across multiple jurisdictions.

IV. Economic Significance of SOGI Data Collection

The economic case for SOGI data collection rests on two facts. First, the scale of disease burden in populations currently uncharacterized by clinical trial datasets is significant, in the multi-trillion-dollar range across the therapeutic areas examined. Second, the operational value of population-complete evidence accrues across drug development, regulatory submission, and post-market surveillance, beyond any single trial.

Pharmaceutical companies that build population-complete data infrastructure see returns in three areas:

- **Stronger clinical insight.** Larger, more representative datasets produce cleaner subgroup analyses. Reliance on post-hoc exploration drops correspondingly.
- **Operational efficiency.** Build the SOGI infrastructure once, apply it across programs. The marginal cost of population-complete capture falls quickly with portfolio scale.
- **Population-relevant therapies.** When the evidence base reflects who actually would receive the treatment, label accuracy and post-market signal detection both improve.

Drivers of Financial Return

- **Therapeutic precision.** More precise data on drug response supports differentiation in competitive markets.
- **Addressable population.** SGM populations represent an estimated 5–10% of the global population, with disproportionate disease burden in specific disease areas.[1,48]
- **Regulatory submission completeness.** Structured demographic capture supports consistency with emerging regulatory recommendations.
- **Corporate reputation.** Pharmaceutical companies recognized for patient-centered approaches attract talent, partnerships, and clinician confidence.
- **Real-world value.** SOGI data extends value beyond clinical trials into post-market surveillance and real-world evidence studies.

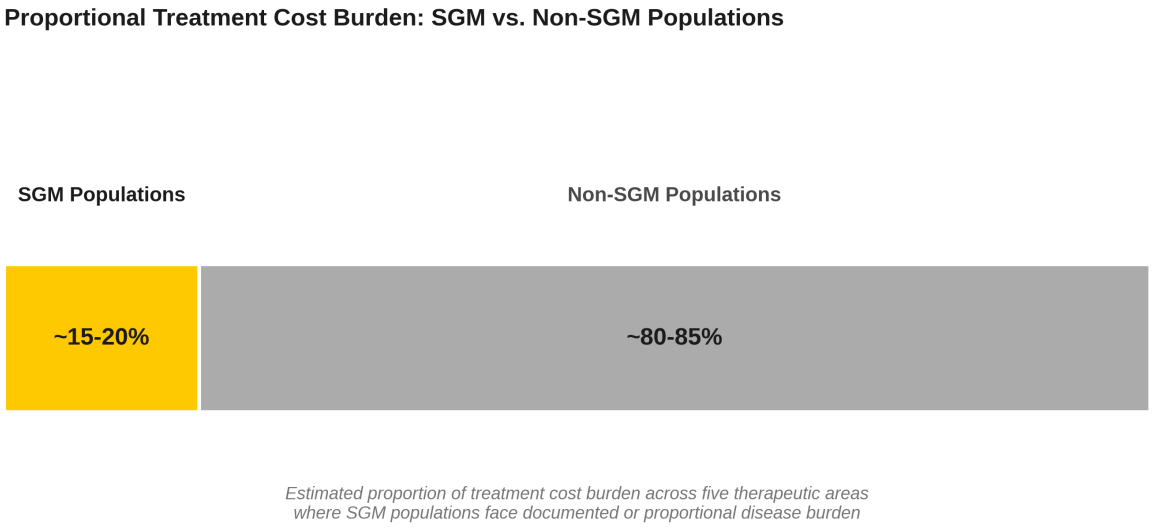

Figure 2. Proportional treatment cost burden across five disease areas where SGM populations face documented or proportional disease burden. See Appendix A for methodology.

Figure 2 illustrates the proportional treatment cost burden across five disease areas where SGM populations face elevated or proportional disease burden, infectious diseases, mental health, hypertension, autoimmune disorders, and type 2 diabetes. The combined treatment cost burden across these therapeutic areas falls in the multi-trillion-dollar range globally. The proportion attributable to SGM populations reflects the scale of treatment cost in populations whose data is largely absent from the trial datasets generating the underlying treatment evidence. The figure represents treatment cost burden, not market opportunity or revenue projection.

More representative data improves the reliability and downstream value of clinical evidence. SOGI data collection supports findings that are relevant, representative, and applicable to the populations who would receive the treatment. This strengthens credibility and reduces risk associated with unrepresentative clinical trial findings.

Case Studies

- **Population-responsive design.** Pharmaceutical companies implementing population-complete trial design, including Merck's requirement of diversity plans for every late-stage clinical trial, its use of census data to prioritize site placement in communities with historically underrepresented populations, and its partnership with the Lazarex Cancer Foundation IMPACT program to remove financial barriers to oncology trial participation, have reported operational efficiencies in enrollment and data quality.
- **Cost of incomplete demographic capture.** Cardiovascular drug development offers the precedent. Women took part in these trials, yet several sex-specific differences in drug response and safety emerged only through post-market surveillance and later label revisions, because the variable that distinguished them was not consistently captured and analyzed. A characteristic that is never recorded cannot be examined, however many members of that population take part.[49]

V. Implementation Pathway

Pharmaceutical companies that adopt voluntary SOGI data collection bring clinical research to all the patients affected by the diseases intended to be treated. The work is operational, not aspirational, protocol language, CRF (case report form) design, eligibility logic, AE (adverse event) coding, SAP (statistical analysis plan) language. What that work could produce:

- Population-complete clinical science, with subgroup analyses that hold up under regulatory and editorial review.
- Patient-centered access strengthened across therapeutic areas where SGM populations have been historically underrepresented.
- Better data supporting financial return, while regulatory and reputational risk are reduced rather than amplified.

This document gives clinical research and healthcare professionals at pharmaceutical companies the consolidated case for SOGI data collection in their organizations. Implementation specifications are set out in SGMA-STD-001 (SOGI Protocol Implementation Standards) and the controlled vocabulary in SGMA-STD-002 (Population Classification Standard).

Applicable SGMA Standards

- **SGMA-STD-001**, SOGI Protocol Implementation Standards. Six-module operational infrastructure: protocol language, eligibility logic, CRF design with SDTM/FHIR mapping, pharmacovigilance, SAP language, and regulatory submission templates.
- **SGMA-STD-002**, Population Classification Standard. Three-layer architecture and controlled vocabulary for population-complete ascertainment, analytical derivation of SGM and non-SGM strata, and key population epidemiologic lenses.

Pharmaceutical Company Engagement

SGMA engages pharmaceutical companies through the Standards Governance and Implementation Committee (SGIC), which provides implementation consultation and conformance review. Pharmaceutical companies may request a structured implementation briefing through sgmastandards.org. SGMA standards are published open-access under Creative Commons Attribution 4.0 and may be implemented directly without licensing.

By addressing the operational requirements, leveraging available partnerships, and aligning with global standards, pharmaceutical companies can take a leading position in building a clinical research ecosystem that reflects the populations served. SGMA welcomes pharmaceutical company engagement on SOGI data collection across global trial programs.

Authorship

Authorship was determined under SGMA-GOV-PP-001 (Publications & Presentations Authorship Standards), which applies the International Committee of Medical Journal Editors (ICMJE) criteria. Named authors meet all four ICMJE criteria: substantial contribution to conception, design, or interpretation; drafting or critical revision for intellectual content; final approval; and accountability for the work.

Authors:

Lauren B. Beach, JD, PhD; Amanda Beasley, PhD, MS; Craig Davenport; Terttu E. E. Haring, MD; Jackie Kent; Garo Kiledjian; Evan Ko; Kelly McKee; Meghan McKenzie, MA; Lisa Moneymaker; Madeline Ormand; Jirair Ratevosian; Donna Sattler; Kemi Williams, MBA; Benjamin Young, MD, PhD, FIDSA.

Authors are listed alphabetically by surname.

Appendix A, Methodology for Figures 1 and 2

Figure 1, Therapeutic Area Distribution

Figure 1 shows the proportional distribution of five disease areas, hypertension, infectious diseases, mental health, autoimmune disorders, and type 2 diabetes, among SGM populations. Proportions are derived by applying documented prevalence differentials from peer-reviewed literature to SGM population estimates, normalized as proportions of total disease burden across the five areas. The figure is illustrative; therapeutic area selection reflects the areas with published prevalence differential data.

Figure 2, Proportional Treatment Cost Burden

Figure 2 illustrates the proportional scale of treatment cost burden across five therapeutic areas where SGM populations face documented or proportional disease burden. The proportion attributable to SGM populations is calculated as follows:

1. Therapeutic area scope.

Five therapeutic areas were selected based on documented or proportional SGM disease burden: infectious diseases, mental health, hypertension, autoimmune disorders, and type 2 diabetes. Combined global treatment cost burden across these areas falls in the multi-trillion-dollar range, drawn from published market research, WHO Global Health Observatory burden-of-disease data, and peer-reviewed disease cost estimates.

2. SGM population baseline.

A baseline SGM population share of approximately 7–10% is applied, derived from multi-country LGBT+ identification data: Ipsos Global Advisor Pride 2024 (26-country survey); Gallup 2024 U.S. LGBTQ+ identification data (9.3%); regional variation across higher-income and lower-and-middle-income studies.[1,48]

3. Therapeutic-area-specific disease burden adjustment.

For therapeutic areas where peer-reviewed literature documents differential prevalence, an area-specific multiplier is applied. Hypertension and infectious diseases are documented at materially higher prevalence in SGM populations relative to non-SGM populations. Mental health prevalence differentials are documented. For therapeutic areas where prevalence differential data is limited (autoimmune disorders, type 2 diabetes), the baseline population share is applied without adjustment.

4. Aggregation.

SGM-attributable treatment cost burden per area = (therapeutic area treatment cost burden) × (SGM population share) × (disease burden multiplier where applicable). Area estimates are summed and presented as a proportion of total treatment cost burden across the five areas.

Interpretive Framework

Figures 1 and 2 represent treatment cost burden across five therapeutic areas where SGM populations face documented or proportional disease burden. They are not revenue projections, pharmaceutical market forecasts, or untapped-market claims. Treatment for individuals within SGM populations is already occurring within each therapeutic area; the figures illustrate the scale of treatment cost burden in populations whose data is not systematically captured in the trial datasets generating treatment evidence.

Known limitations:

- Baseline SGM population share derives from self-reported identification surveys that vary by country, methodology, and survey year.
- Disease burden multipliers are applied only where peer-reviewed data exists; areas with limited differential data may be under- or over-weighted.
- Therapeutic area treatment cost sizing depends on source methodology, which varies across published market research.

- These are top-down illustrative estimates. A peer-reviewed bottom-up methodology is in development with academic partners and will supersede these estimates upon publication.

References

References below have been verified for current accessibility (April 2026). Where v1.0 references could not be verified or had been superseded by changes in federal data collection (per Williams Institute 2026), substitutions or removals were applied; see editorial change log accompanying this document.

1. Jones JM. LGBTQ+ Identification in U.S. Rises to 9.3%. Gallup. 2025. https://news.gallup.com/poll/656708/lgbtq-identification-rises.aspx

2. Deckman M, Jones PR, Orcés D, et al. A Political and Cultural Glimpse Into America's Future: Generation Z's Views on Generational Change and the Challenges and Opportunities Ahead. PRRI. 2024.

3. Quinn GP, Pratt-Chapman ML, Meersman SC, et al. Barriers and facilitators to sexual orientation and gender identity (SOGI) data collection. J Clin Oncol. May 2021;39(15). doi:10.1200/JCO.2021.39.15_suppl.e18520

4. Studna A. Complexities in Data Collection. Applied Clinical Trials. 2024.

5. Markey N, Howitt B, El-Mansouri I, et al. Clinical trials are becoming more complex: a machine learning analysis of data from over 16,000 trials. Sci Rep. 2024;14:3514. doi:10.1038/s41598-024-53211-z

6. CDISC. Sexual Orientation & Gender Identity (SOGI). https://www.cdisc.org/kb/ecrf/sexual-orientation-gender-identity-sogi

7. Brown J, Cunningham M, Coney JM, et al. Elevatum study design and rationale: a phase 4 trial of faricimab in underrepresented patients with DME. Invest Ophthalmol Vis Sci. Jun 2024;65(7).

8. Diversity Action Plans to Improve Enrollment of Participants from Underrepresented Populations in Clinical Studies. U.S. Food and Drug Administration. June 2024 (draft). Restored to FDA website by court order, February 2025.

9. Food and Drug Omnibus Reform Act of 2022 (FDORA), Section 3601 (codified at 21 U.S.C. § 355(z) and § 360j(g)).

10. Executive Order 14168 (January 27, 2025). Executive Office of the President.

11. Jennifer Kates UR, Beamesderfer A, Salganicoff A, Dawson L. Health and Access to Care and Coverage for Lesbian, Gay, Bisexual, and Transgender (LGBT) Individuals in the U.S. KFF. 2018.

12. Hill-Briggs F, Adler NE, Berkowitz SA, et al. Social Determinants of Health and Diabetes: A Scientific Review. Diabetes Care. Nov 2020;44(1):258-79. doi:10.2337/dci20-0053

13. CDC. Non-Medical Factors that Affect Alzheimer's Disease and Related Dementias Risk. 2024.

14. Curry CW, Felt D, Beach LB, et al. Lifetime Asthma Prevalence and Correlates Among US Youths by Sexual Identity and Race/Ethnicity, 2009-2017. Am J Public Health. May 2020;110(7):e1-e8. doi:10.2105/AJPH.2020.305664

15. American Lung Association. Current Asthma Demographics.

16. Katon JG, Plowden TC, Marsh EE. Racial disparities in uterine fibroids and endometriosis: a systematic review. Fertil Steril. Mar 2023;119(3):355-363. doi:10.1016/j.fertnstert.2023.01.022

17. Mitro SD, Dyer W, Lee C, et al. Uterine Fibroid Diagnosis by Race and Ethnicity in an Integrated Health Care System. JAMA Netw Open. Apr 2025;8(4):e255235. doi:10.1001/jamanetworkopen.2025.5235

18. Shanker A, Korniejczuk M. Collecting Data on Sexual Orientation and Gender Identity. Public Health Rep. 2024. doi:10.1177/00333549241308401

19. Tran NK, Rosendale N, Lunn MR. Advancing Data Collection of Sexual Orientation and Gender Identity in Cardiology. JAMA Cardiol. Mar 2024;9(3):206-208. doi:10.1001/jamacardio.2023.5264

20. Goldhammer H, Grasso C, Katz-Wise SL, et al. Pediatric sexual orientation and gender identity data collection in the electronic health record. J Am Med Inform Assoc. Jun 2022;29(7):1303-1309. doi:10.1093/jamia/ocac048

21. Beach LB, Flentje A, Freeman JB. In order to count, lesbian, gay, bisexual, transgender, queer, intersex, and asexual people must be counted. Health Psychol. Mar 2025;44(3):332-334. doi:10.1037/hea0001478

22. United Nations Department of Economic and Social Affairs, Population Division. World Population Prospects.

23. World Health Organization. Mental health: strengthening our response. 2022.

24. Ipsos. LGBT+ Pride 2024: A 26-Country Ipsos Global Advisor Survey. 2024.

25. Gao S, Bell EC, Zhang Y, Liang D. Racial Disparity in Drug Disposition in the Digestive Tract. Int J Mol Sci. Jan 2021;22(3). doi:10.3390/ijms22031038

26. Jordan IK, Sharma S, Nagar SD, Marino-Ramirez L. The Apportionment of Pharmacogenomic Variation: Race, Ethnicity, and Adverse Drug Reactions. Med Res Arch. Sep 2022;10(9). doi:10.18103/mra.v10i9.2986

27. Taliun S. Short guideline on the use and reporting of race/ethnicity/ancestry. International Genetic Epidemiology Society.

28. Use of Population Descriptors in Genomics. National Human Genome Research Institute. 2024.

29. National Human Genome Research Institute. Pharmacogenomics fact sheet.

30. Jackson CL, Agenor M, Johnson DA, Austin SB, Kawachi I. Sexual orientation identity disparities in health behaviors, outcomes, and services use. BMC Public Health. Aug 2016;16(1):807. doi:10.1186/s12889-016-3467-1

31. Shover CL, DeVost MA, Beymer MR, et al. Using Sexual Orientation and Gender Identity to Monitor Disparities in HIV, STIs, and Viral Hepatitis. Am J Public Health. Nov 2018;108(S4):S277-S283. doi:10.2105/AJPH.2018.304751

32. Kamen CS, Pratt-Chapman ML, Meersman SC, et al. SOGI Data Collection in Oncology Practice: ASCO Survey. JCO Oncol Pract. Aug 2022;18(8):e1297-e1305. doi:10.1200/OP.22.00084

33. Alpert AB, Obedin-Maliver J, Gjelsvik A, et al. Does cancer clinical trial enrollment for sexual and gender minority people differ from heterosexual, cisgender people? Contemp Clin Trials. Nov 2024;146:107695. doi:10.1016/j.cct.2024.107695

34. Sorensen JL, West MM, Racila AM, et al. Challenges in collecting information on sexual orientation and gender identity for cancer patients. Cancer Causes Control. Nov 2024;35(11):1433-1445. doi:10.1007/s10552-024-01897-x

35. Pratt-Chapman ML, Miech EJ, Mullins MA, et al. Difference-Makers for Collecting SOGI Data in Oncology Settings. Cancer Med. Mar 2025;14(5):e70727. doi:10.1002/cam4.70727

36. Scout N. Unique Issues Facing Sexual and Gender Minorities in Cancer. Cancer Discov. Jun 2023;13(6):1297-1300. doi:10.1158/2159-8290.CD-23-0455

37. Meyer IH, Bouton LJ. Removal of Sexual Orientation and Gender Identity from Federal Data Collections. Williams Institute, UCLA School of Law. February 2026.

38. World Medical Association. Declaration of Helsinki. https://www.wma.net/policies-post/wma-declaration-of-helsinki/ See also: International Council for Harmonisation. ICH E6: Good Clinical Practice. https://www.ich.org/page/efficacy-guidelines

39. Kuri L, Sattler D, Brescher J. BMS and the LGBTQ+ Community Improve Clinical Data to Achieve Health Equity. The Conference Forum: DPHARM. https://theconferenceforum.org/editorial/bms-and-the-lgbtq-community-improve-clinical-data-to-achieve-health-equity

40. Genentech (sponsor). Sexual orientation and gender identity data: An observational study assessing the feasibility of SOGI collection in clinical research and patient assistance programs. PLOS One. 2025. doi:10.1371/journal.pone.0332805

41. Brown J, Cunningham M, Coney JM, et al. Elevatum study design. Invest Ophthalmol Vis Sci. Jun 2024;65(7).

42. GSK. Representative Clinical Studies. https://www.gsk.com/en-gb/innovation/trials/representative-clinical-studies/

43. Azad A, Musto C, Netanel S, et al. LIBERTAS: A degendered and transgender-inclusive phase 3 study. J Clin Oncol. Jun 2024;42(16).

44. Streed C Jr, Navarra M, Beach LB, et al. Sex, sexual orientation, and gender identity data collection across electronic health record platforms. JAMIA Open. Nov 2024;7(4):ooae127. doi:10.1093/jamiaopen/ooae127

45. Streed C. Sexual Orientation and Gender Identity Data Collection. AMA Center for Health Equity.

46. U.S. Food and Drug Administration. Evaluation of Sex- and Gender-Specific Data in Medical Device Clinical Studies.

47. National Academies of Sciences, Engineering, and Medicine. Measuring Sex, Gender Identity, and Sexual Orientation. National Academies Press. 2022.

48. Ipsos LGBT+ Pride 2024 Survey, op. cit.

49. Michos ED, Reddy TK, Gulati M, et al. Improving the enrollment of women and racially/ethnically diverse populations in cardiovascular clinical trials. Am J Prev Cardiol. Dec 2021;8:100250. doi:10.1016/j.ajpc.2021.100250