



Can We Learn from the “Wisdom of the Crowd”? Finding the Sample-Size Sweet Spot - an Analysis of Internet-Based Crowdsourced Surveys of Fertility Professionals

Gon Shoham^{1*}, Milton Leong², Ariel Weissman³ and Yuval Yaron⁴

*Correspondence: gonshoha@tauex.tau.ac.il



CrossMark

← Click for updates

¹Sackler Faculty of Medicine, Tel Aviv University, P.O.B. 39040, Ramat Aviv, Tel Aviv 69978, Israel.

²IVF Clinic, The Women's Clinic, 12/F, Central Tower, 28 Queen's Road Central, Central, Hong Kong, China.

³IVF Unit, Department of Obstetrics & Gynecology, Edith Wolfson Medical Center, 62 Ha-Lokhamim St., Holon 5822012, Israel; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

⁴Prenatal Genetic Diagnosis Unit, Genetics Institute, Tel Aviv Sourasky Medical Center, 6 Weizmann Blvd., Tel Aviv, Israel; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Abstract

Background: The purpose of this research was to calculate the minimum sample size needed to obtain reliable results from crowdsourced retrospective online surveys of IVF clinics, where the sample was IVF cycles performed annually per clinic— a new metric that may offer more survey flexibility than number of clinics or respondents.

Methods: This analysis used two statistical formulas to calculate sample sizes and confidence intervals, initially assessing a published self-report survey conducted online by IVF-Worldwide of global IVF practices. The survey covered 592,900 IVF cycles from 795 clinics worldwide. A subset of 275,600 European IVF cycles was used as a test sample population, which was compared statistically with the actual survey population. To validate results, two additional geographic subsets (North America and Asia) from the initial survey and three additional previously published surveys were also evaluated in same fashion. Only one survey entry was accepted per clinic.

Results: Results showed that to obtain reliable survey outcomes, the estimated minimum sample size was 35,000 cycles. Given a 99% confidence level, 0.5 probability, and a 5% estimation error, the minimum sample size for Europe was 35,340. Similarly, sample sizes for earlier surveys and other continent subsets were between 35,280 and 35,340.

Conclusions: Surveyors often strive for small sample sizes to save cost, effort, time and/or management overhead. Currently, however, sample size standards for online surveys do not exist. The results presented here suggest that sample sizes below 35,000 may lead to unreliable results. Finding the right sample size for multiple-choice crowdsourced surveys will save cost and effort, accelerate research, and empower surveys originally deemed too difficult and/or expensive due to inordinate sample sizes required. Surveys that count number of IVF cycles as the survey population may offer researchers more options for how to conduct and analyze surveys. Further research is required to apply findings to other areas of medicine, such as surveying using the number of procedures performed by clinic.

Keywords: Infertility, survey, sample size, statistics, crowdsource, IVF-Worldwide

Introduction

The *wisdom of the crowd* was described by Aristotle in *Politics* [1] as the collective opinion of a group of individuals rather than of a single expert. This phenomenon evolved over time

from the classic wisdom-of-the-crowd finding involving point estimations of a continuous quantity [2] to applications in fields as diverse as politics, economics, business strategy, and more [3]. In medicine, many studies have tested this theory by comparing

diagnostics and decision-making disciplines such as radiology [4], dermatology, and breast cancer [5].

What represents a large enough crowd has always been of concern, as many research studies using this study format were under powered because of sample sizes that were too small or insufficient data. Small sample size is one of the frequently mentioned limitations in many published infertility studies, driving recommendations that further research be conducted involving more patients, more case studies or more IVF cycles [6-8]. Even studies involving thousands of IVF cycles, at times, are still considered underpowered. Clinical studies across all disciplines have demonstrated positive correlations between large numbers of procedures and statistically significant outcomes.

The advent of internet-based surveys that leverage the networked world, has influenced and penetrated almost every aspect of medical research. Indeed, internet surveying has been suggested as a new mainstream method of performing medical research [9]. The next evolutionary stage in online surveying was *crowdsourcing*, with crucial prerequisites being open call formats and large networks [10]. Medical researchers have leveraged crowdsourcing to help conduct important research, with numerous studies published across the spectrum of peer-reviewed journals.

Crowdsourcing can potentially improve research project quality, cost, and speed while engaging large segments of the public [11]. Its results can be used to create new research tools and techniques, discover medical insights, and solicit innovative ideas for novel science. Crowdsourcing is a potential alternative for studies that require or could benefit from analyzing large amounts of data, automating manual data processing, surveying subjects worldwide, tapping into specific skills from diverse members of the public, soliciting global subject matter experts or accessing diverse subject pools at lowcost [11].

With ever-increasing interest in data mining and machine learning, too much data can cause information overload. This stands in contrast to the “more is better” data collection approach that is the current standard in most medical publishing. Moreover, machine-learning algorithms do not work well in borderline cases of medical uncertainties [12]. In some of these instances, the wisdom-of-the-crowd strategy may help counter this problem by channeling a vast number of opinions into one conclusion. Therefore, the question arises: when is data considered enough? There should be a balance for survey size in which there is enough data to represent a crowd’s opinion, while the addition of more responses would not necessarily further improve outcome reliability for a defined confidence level. Existing statistical sampling theory does not provide insights regarding sample size that can be applied across multiple self-report studies, specifically minimum sample sizes for large datasets such as IVF cycles.

This aim of this paper was to determine the minimum sample size for a new metric for survey sample size that would deliver high reliability while reducing the number of physician/

clinic responses required. The minimum sample size of the new metric “IVF cycles performed annually per clinic” was successfully calculated, achieving the aims of the research.

Materials and methods

Research usually involves an analysis of the relative benefit of an active treatment over a control. To measure this benefit, statistical tools such as relative risk, relative risk reduction or odds ratios are utilized in clinical and epidemiological investigations. For clinical decision-making, studies use the measure “numbers needed to treat” (NNT) which conveys both statistical and clinical significance [13].

In contrast to studies that assess *outcomes* of treatments, surveys often examine *choices* of treatments made by medical professionals based on certain scenarios, but not necessarily linked to study evidence. In these cases, we look for the common *opinion and practice* of clinicians and researchers, and NNT is not relevant. Sample size considerations for these kinds of surveys are not based on estimations of treatment effects, but rather on precisely estimating a population proportion.

The IVF-Worldwide website (IVF-worldwide.com) [14] membership is a vast global community of reproductive-health physicians and researchers. Numerous surveys are conducted on the site, integrating input from hundreds of medical units (IVF clinics and obstetrics practices), representing hundreds of physician- and researcher-participants, and reporting results based on hundreds of thousands of in vitro fertilization (IVF) cycles from geographic regions all over the world.

Data and findings from previously performed IVF-Worldwide surveys [15] were used as the core of this study. Using IVF cycles instead of individual clinics as the sample population expanded the sample size metric from hundreds of clinics to hundreds of thousands of cycles. Therefore, this dataset allowed the study researchers to use multiple self-report surveys to draw general conclusions. For example, it enabled the researchers to: (i) calculate the minimum survey sample size(n) needed to produce reliable results from the crowd sourcing surveys, and (ii) develop statistical measures and tools to help optimize research study effectiveness without requiring large, unwieldy survey data sets.

The study harnessed IVF-Worldwide’s physician and researcher surveys and results that have quantified the popularity of treatment protocols used in fertility clinics worldwide. In coordination with IVF-Worldwide, survey results were generated from these unprecedented amounts of crowdsourced data. These surveys did not replace evidence-based-medicine and did not suggest that clinicians follow reported practices. They, however, reflected the opinions and experience of medical directors from hundreds of IVF units worldwide (one representative response per unit).

We used the IVF-Worldwide.com survey database for source survey data. See **Figure 1** for the PRISMA Flow Diagram for survey selection. For statistical calculations, we used two common statistical tools for survey analysis [16], which indicate how well

a sample statistic estimates the underlying population value. The sample size formula was calculated as follows: sample size,

$$= \frac{Z^2 * p(1-p)}{e^2}$$
 where Z was the z-score, p was the sample proportion, and e was the margin of error. The confidence interval (C/I) formula was calculated as follows:

$$C/I = \bar{x} \pm Z \frac{s}{\sqrt{n}}$$
 where x was the mean, s was the standard deviation, and n was the sample size.

The statistical calculations were made on the data from the largest survey performed: "Anti-Müllerian hormone (AMH) and antral follicular count (AFC)" [17]. As a starting point, we took a subset of results from a single geographic area – survey responses from all European respondents – to help ensure high coherence in respondent education, the pool of medical protocols and practices used. We also chose Europe because the continent leads the world in the number of assisted reproductive technology (ART) procedures performed, representing approximately 50% of all reported treatment cycles [18]. We

used the sample size and C/I formulas to predict the minimum number of IVF cycles that would ensure that the results stayed within the 5% median estimation error (e) while keeping a 99% confidence level (Z=2.58). We used the worst-case scenario in which the probability of each answer was 0.5 (p=0.5). We also defined the true results to be the final global survey results, comparing all subset results to them.

After applying the formulas to find the minimum number of IVF cycles that met the criteria, we tested the theory on three additional published surveys of IVF professionals with large sample sizes, again, isolating the European respondents. The surveys used were: "Egg collection and embryo transfer techniques" [19], "The use of gonadotropins and biosimilars in assisted reproductive technology (ART) treated cycles" [20], and "Preimplantation genetic screening (PGS): what is my opinion?" [21]. We also verified IVF cycle sample size results against other geographic regions in the original survey. In each calculation, we made sure that the unit size distribution was similar to the global sample. All samples had an absolute average distribution difference of less than 3.5%.

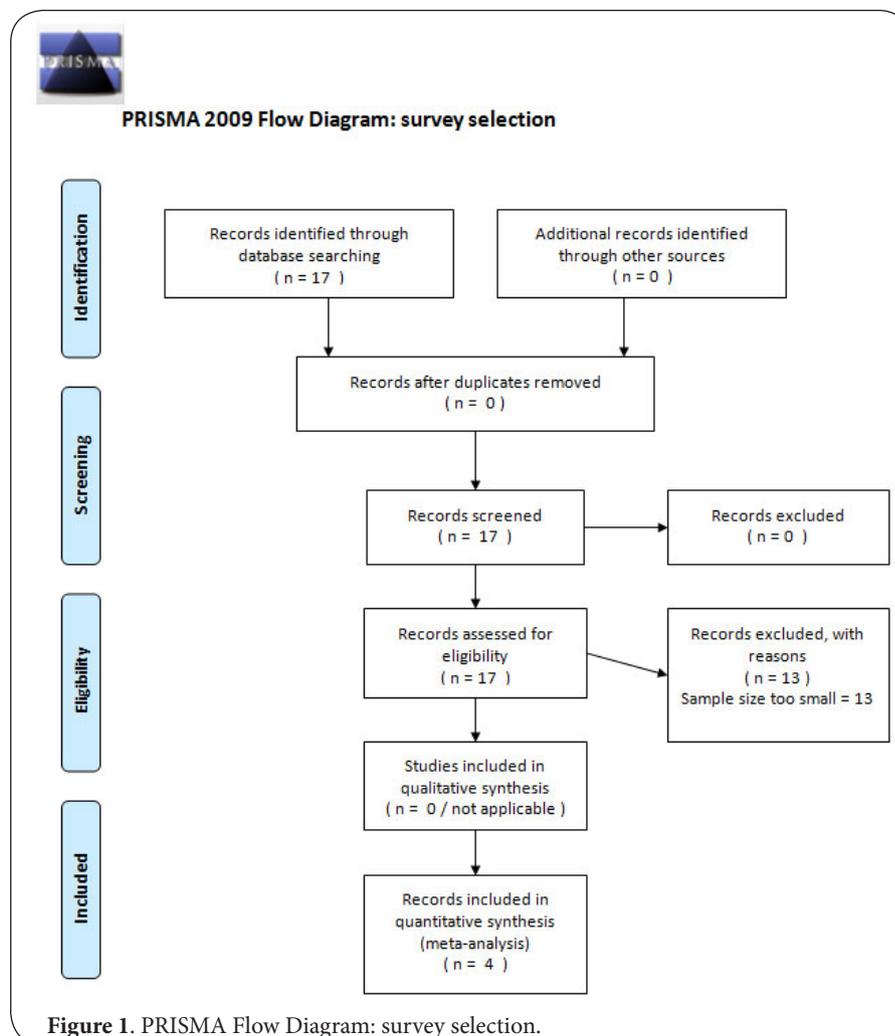


Figure 1. PRISMA Flow Diagram: survey selection.

Quality assurance

To minimize duplicate clinical unit survey reports and eliminate possible false data that would double-count or skew results, we limited responses to one per clinic. We used a software program (BF Survey, Tamlyn Software, Sydney, NSW, Australia) that compared the consistency of three parameters from self-reported survey-unit data with existing unit data from the IVF-Worldwide website, as previously described [22]. These parameters included the unit name, country, and e-mail address. At least two parameters had to match between the survey and the website for clinical unit data to be included in the study. If two survey responses shared at least two parameters, the duplicate survey results with the later date were discarded.

Statistical analysis

The analysis was based on the number of annual IVF cycles reported, and not on the number of units or respondents in the study. The number of IVF cycles that a unit could report was between 100 and 4500, limiting the contribution of a single unit to a maximum of 0.76% (4500/592,900). Thus, the relative proportion of each unit's answers reflected the total proportion of IVF cycles performed annually, representing the knowledge and experience of the physicians in the unit.

Each survey was structured as a series of multiple-choice questions in which respondents selected a single answer for most questions. (There was a small number of questions in which multiple answers could be given). Results were calculated by using the formulas described in previously published research from the IVF-Worldwide network [21]. For example, for a single-answer or multiple-answer question with four choices (a, b, c, d), the following results were calculated to determine the percentage of responses:

$$'a' \% = \frac{\sum \text{Number of cycles of units who answered 'a'}}{\sum \text{Number of cycles of all responding survey units}} * 100$$

$$'b' \% = \frac{\sum \text{Number of cycles of units who answered 'b'}}{\sum \text{Number of cycles of all responding survey units}} * 100$$

$$'c' \% = \frac{\sum \text{Number of cycles of units who answered 'c'}}{\sum \text{Number of cycles of all responding survey units}} * 100$$

$$'d' \% = \frac{\sum \text{Number of cycles of units who answered 'd'}}{\sum \text{Number of cycles of all responding survey units}} * 100$$

Compliance with ethical requirements

IVF-Worldwide surveys were self-reports on activity volumes and opinions on medical practices from physicians and researchers. The surveys conducted and this study did not involve research on human or animal subjects; therefore, institutional review board approval was not required. Surveys were conducted as open-access questionnaires to IVF-Worldwide.com members who voluntarily answered the study questions. Data collected for this research was anonymous; no patient's details were required.

Results

In the "Anti-Müllerian hormone (AMH) and antral follicular count (AFC)" survey, the sample size was 592,900 IVF cycles (Table 1). We studied the European geographical subset of the complete survey set, which accounted for 275,600 IVF cycles. Using iterations, we created new survey results, cutting the sample size in half for each iteration, until we first received a mean error margin of 5% or lower. The error was calculated as the average absolute difference of each answer compared to the answers for the 275,600 IVF cycles. We used spreadsheet software (Microsoft Office Excel, California, USA) to select the samples and analyze results. We avoided selection bias by applying the Excel "random" function to produce random sample populations. We repeated this process five times for each sample size. Results presented were the average of these five selections. In addition, we made sure that, for each iteration, the unit-size distribution would be similar to the global sample; therefore, all samples created demonstrate an absolute average of less than 3.5% distribution difference.

We found that at about 35,000 (exactly 35,340) cycles, the mean estimation error dropped below 5%. To estimate errors of 1% and 5%, we created a C/I table with a confidence level of 99% and a probability of 0.5. We first recalculated the formula for a 5% error using 35,000 as the sample size. Then we recalculated the formula again, based on the 5% parameters, to see if the error margin for using 275,600 IVF cycles as the sample size would be smaller than 1%. This defined 275,600 cycles as the best case for which the estimation error would

Table 1. The Application of Revised Sample Sizes to IVF-Worldwide Survey Results.

Survey	Total IVF cycles	Total IVF units	Total European IVF cycles	Total European units	European IVF cycles (avg.) for 5% estimation error	European units (avg.) for 5% estimation error	European units distribution (avg.) for 5% estimation error
Anti-Mullerian hormone and antral follicular count	592,900	795	275,600	329	35,340	43	2.2
Gonadotropins	218,300	314	97,000	135	35,320	50	2.2
Egg collection and ET	258,300	355	115,300	143	35,300	52	2.1
PGS (Yes only)	263,100	271	91,600	70	35,280	28	2.3

be 1%. Approximately 35,000 IVF cycles provided accurate results, and this IVF-cycle sample size was relatively easy to obtain. These results were displayed in **Figure 2**.

The next step was to validate the findings by comparing the ~35,000-cycle sample-size results against results from additional surveys. We chose three published surveys that had input from about 100,000 cycles in Europe (**Table 2**). In all three, we compared the results for approximately 35,000 cycles to the results of the complete European set (between 35,280 and 35,340 cycles were used, so as not to split a single unit's response). We used the Excel random function to create five different sample results for each survey, and calculated the absolute difference between each answer and the average mean error of the new minimum sample size (**Figure 3**).

The results showed that for the "Egg collection and embryo transfer techniques" and "The use of gonadotropins and bio-similars in ART treated cycles" surveys, the mean error was below 4.5%, and in the "Preimplantation genetic screening (PGS): what is my opinion?" survey, the mean error was 5%. Moreover, in all surveys, the 75th percentile of each answer's absolute difference was within a 7% mean error, which increases the confidence in our claim.

To strengthen our results and address selection biases, we further analyzed the "Anti-Mullerian hormone and antral follicular count" survey by comparing the ~35,000 cycle results to the complete survey results from other continents. We chose to analyze North America (USA and Canada) and Asia, the continents with the second and third most IVF cycles, respectively. The results (**Figure 4**) were consistent with the results from the European cycle comparison, with a mean error estimation of 3.1% (USA & Canada) and 5% (Asia). These

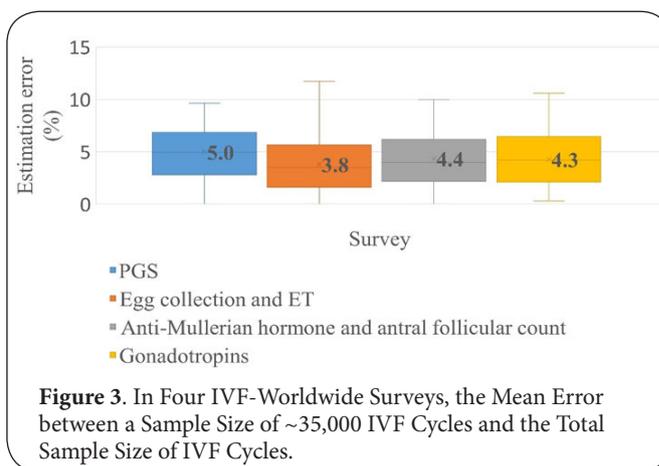


Figure 3. In Four IVF-Worldwide Surveys, the Mean Error between a Sample Size of ~35,000 IVF Cycles and the Total Sample Size of IVF Cycles.

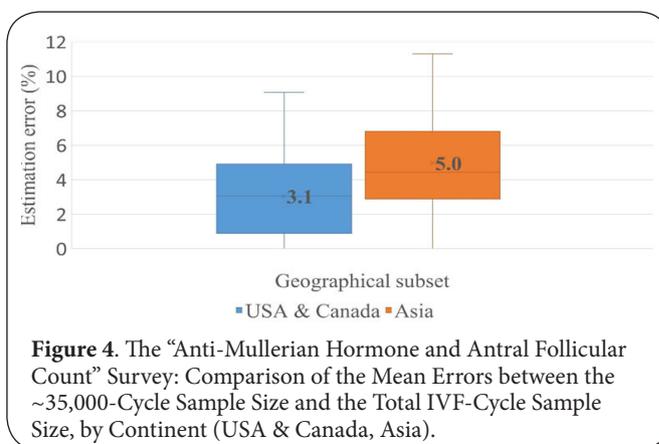


Figure 4. The "Anti-Mullerian Hormone and Antral Follicular Count" Survey: Comparison of the Mean Errors between the ~35,000-Cycle Sample Size and the Total IVF-Cycle Sample Size, by Continent (USA & Canada, Asia).

results showed that a ~35,000 cycle sample size gave reliable results regardless of region and topic surveyed.

Discussion

IVF-Worldwide surveys seek to obtain insights from researchers and physicians worldwide on common medical practices and beliefs, based on their vast combined experience and diverse education. The survey answers reflected hundreds of thousands of self-reported cases, and the survey findings showed that "wisdom of the crowd" surveys have certain advantages. The natural tendency to seek larger population sizes for medical trials is a reasonable desire since a large number of case studies usually reflects a wider range of opinions and responses to treatments, which may help discover findings not otherwise seen in small sample sizes, such as adverse effects. Using large sample sizes enables researchers to reach

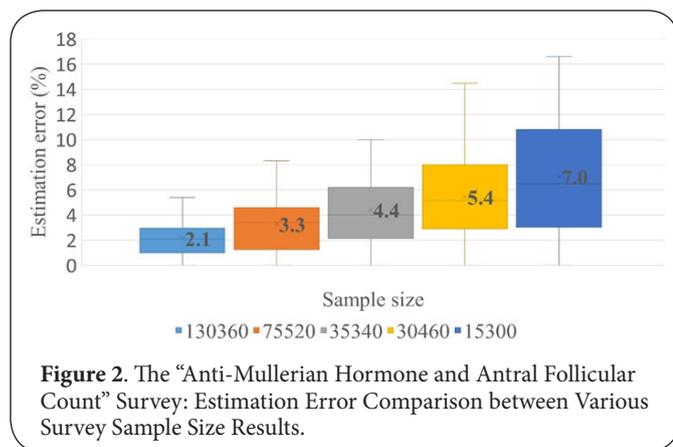


Figure 2. The "Anti-Mullerian Hormone and Antral Follicular Count" Survey: Estimation Error Comparison between Various Survey Sample Size Results.

Table 2. The "Anti-Mullerian Hormone and Antral Follicular Count" Survey: Original and Revised Survey Size and Number of Units Surveyed for North America and Asia.

Survey	Total IVF cycles	Total IVF units	Comparison IVF cycles (avg.)	IVF units (avg.) for comparison IVF cycles (avg.)	IVF units distribution (avg.), for comparison IVF cycles (avg.)
USA & Canada	64,800	121	35,260	63	3.1
Asia	140,300	174	35,080	47	5

conclusions with better confidence and certainty. However, obtaining larger research sample sizes may not be practical or possible; they require more financial resources and may increase the survey duration. These factors can discourage research and prolong result acquisition and analysis, by which time the findings may have been published elsewhere, making the research redundant and outdated. A balance between these conflicting interests may help minimize cost and effort while keeping research current.

Most IVF self-report surveys report findings with reference to clinic and/or physician activity or opinion. They do not use IVF cycles as the sample population. The international scope and high response rates of the IVF-Worldwide surveys analyzed produced an extremely large sample size, which is rare in this discipline. In addition, an IVF-cycle-based survey (versus a clinic-based or physician-based survey) may help level the playing field slightly between large and small clinics.

Each future IVF survey conducted needs to perform its own statistical calculations to determine the appropriate sample size. However, this study gives researchers a new dimension by which to arrive at statistically sound, reliable results while processing fewer physician or clinic responses. For example, given that the average clinic in the initial study performed 745 cycles (592,900 cycles/795 clinics), an estimated 47 clinics (35,000/745) would give researchers 35,000 cycles.

Until now, researchers may have chosen not to embark on large-scale surveys due to the number of clinics or physicians that may have been required. Obtaining a sample size of 35,000 can potentially save surveyor time and effort, as well as survey and analysis expense, while still giving surveyors decisive analysis results. Therefore, requiring a sample size of only 35,000 can empower researchers to perform surveys that were originally deemed impossible.

In search for a statistical tool similar to NNT that struck a balance among benefits of crowdsourcing, data mining, and information overload, we adopted the survey size and C/I formulas used in opinion surveys. One drawback of this method was that, unlike surveys of complete populations, such as election surveys, we did not have the true results with which to compare the survey results. However, the European survey results themselves were based on 275,600 IVF cycles. And according to the European Society of Human Reproduction and Embryology [18], in 2016, the latest year for which figures were available, 800,000 treatment cycles were performed. Since the number of cycles in the 2014 survey would not likely have been different in orders of magnitude, the survey covered approximately 35% (275,600/800,000) of the cycles in Europe. Statistical sampling theory defines that the mean error decreases when the sample size grows. Therefore, any difference between the survey's final results and the true results would be negligible. This conclusion supported our claim that web surveys like those conducted by IVF-Worldwide.com can predict common beliefs and practices since the differences in mean error estimation decrease as sample sizes increase. Therefore,

in the surveys we analyzed, we can claim with certainty that the answers reflected unequivocal trends, giving the results a very high level of reliability in the field of infertility.

Limitations

The originally published survey results should be used with great caution. By no means should they replace evidence-based-medicine and did not suggest that clinicians follow the practices supported in surveys. They simply reflected the opinions and experience of medical directors from hundreds of clinics worldwide at a specific point in time. These findings did raise the possibility that this "wise crowd" may make different treatment choices than those recommended in the literature and published as evidence-based medicine. Therefore, one must question why. Survey findings may be a wake-up call and catalyst for further research to discover why clinicians do not practice according to clinical evidence. This, in turn, may contribute to a roadmap toward resolving discrepancies. In the uncommon case in which a single clinic submitted multiple entries, only the first survey response was accepted. This implies that the surveys did not assess situations in which opinions differed within each clinic. Data on IVF cycles likely correlated with data on IVF clinics. However, this correlation was not evaluated in these analyses since this could introduce bias into the calculations. Such assessments should be considered in future research. The objective of this research was to determine a new metric and minimum sample size, and the intent was not solely to optimize efficiency. This approach does not replace the need for each survey to calculate sample size to prevent sampling errors. If statistical calculations recommend a survey size of over 35,000, then surveyors should use the larger sample size that was calculated. The paper does not apply the results to research that was not survey oriented, such as actual procedures performed. Given that a relatively small number of clinics would be needed to produce 35,000 IVF cycles, researchers should be cautious about the international distribution of the responses and applying findings to international populations.

Conclusions

Our calculations produced the number of IVF cycles that delivers results with both a high confidence level and a reasonably low mean estimation error, with statistical significance similar to results from a survey with more than five times the sample size. As personalized treatments continue to develop worldwide, and medical databases grow, new treatment protocol creation will rely more heavily on algorithms that consider enormous numbers of cases gathered from physicians, such as IVF-Worldwide.com surveys. This massive amount of data will create the right conditions for evaluating new protocols or updating existing protocols by revealing new correlations and discoveries. However, bigger is not always better, and the results or conclusions drawn from larger studies may not always improve or may improve only marginally after reach-

ing a threshold sample size. For IVF, we believed that reaching ~35,000 cycles was a realistic study sample goal, allowing study results and assumptions to be applied to complete representative populations. To strengthen this claim, further research is required. We expect that future surveys will include more participants, meaning more IVF cycles. Therefore, we will be able to test our theory against true results in future surveys. Finally, this research can spark interest across a spectrum of medical fields, encouraging other researchers to test and potentially apply this method to their disciplines.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Authors' contributions	GS	ML	AW	YY
Research concept and design	✓	✓	✓	✓
Collection and/or assembly of data	✓	✓	✓	✓
Data analysis and interpretation	✓	✓	✓	✓
Writing the article	✓	✓	✓	✓
Critical revision of the article	✓	✓	✓	✓
Final approval of article	✓	✓	✓	✓
Statistical analysis	✓	✓	✓	✓

Acknowledgements

The researchers would like to thank Norbert Gleicher, MD, from the center for Human Reproduction, New York, New York, USA, for his insight, and the hundreds IVF units who took the time and effort to complete the surveys posted on IVF-Worldwide.com.

Publication history

EIC: Jimmy Efir, East Carolina University, USA.
 Received: 14-Aug-2019 Final Revised: 22-Sept-2019
 Accepted: 24-Sept-2019 Published: 30-Sept-2019

References

- Ober J. **An Aristotelian middle way.** 2009. | [Pdf](#)
- Galton F. **Vox populi.** *Nature.* 1907; **75**:450-451. | [Pdf](#)
- Rich L. **Tapping the Wisdom of the Crowd.** *The New York Times.* 2010. | [Article](#)
- Wolf M, Krause J, Carney PA, Bogart A and Kurvers RH. **Collective intelligence meets medical decision-making: the collective outperforms the best radiologist.** *PLoS One.* 2015; **10**:e0134269. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
- Kurvers RH, Herzog SM, Hertwig R, Krause J, Carney PA, Bogart A, Argenziano G, Zalaudek I and Wolf M. **Boosting medical diagnostics by pooling independent judgments.** *Proc Natl Acad Sci U S A.* 2016; **113**:8777-82. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
- Schieve LA, Meikle SF, Ferre C, Peterson HB, Jeng G and Wilcox LS. **Low and very low birth weight in infants conceived with use of assisted reproductive technology.** *N Engl J Med.* 2002; **346**:731-7. | [Article](#) | [PubMed](#)
- Pandian Z, Gibreel A and Bhattacharya S. **In vitro fertilisation for unexplained subfertility.** *Cochrane Database Syst Rev.* 2015; CD003357. | [Article](#) | [PubMed](#)
- Weiss NS, Nahuis M, Bayram N, Mol BW, Van der Veen F and van Wely M. **Gonadotrophins for ovulation induction in women with polycystic ovarian syndrome.** *Cochrane Database Syst Rev.* 2015; CD010290. | [Article](#) | [PubMed](#)
- Schonlau M. **Will Web Surveys Ever Become Part of Mainstream**

- Research? *J Med Internet Res.* 2004; **6**. | [Article](#)
- Howe J. **The Rise of Crowdsourcing.** *Wired.* 2006; **14**:1-4. | [Article](#)
- Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, Seymour AK and Merchant RM. **Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review.** *J Gen Intern Med.* 2014; **29**:187-203. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
- Chen JH and Asch SM. **Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations.** *N Engl J Med.* 2017; **376**:2507-2509. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
- Cook RJ and Sackett DL. **The number needed to treat: a clinically useful measure of treatment effect.** *BMJ.* 1995; **310**:452-4. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
- IVF-Worldwide. 2010. | [Website](#)
- IVF-Worldwide Surveys. *IVF-Worldwide.* 2010. | [Article](#)
- Calder K. **Statistical inference.** 1953. New York, NY, USA: Holt.
- Anti-Mullerian hormone (AMH) and antral follicular count (AFC) Survey. *IVF-Worldwide.* 2014. | [Article](#)
- ART fact sheet. **European Society of Human Reproduction and Embryology (ESHRE).** 2018. | [Website](#)
- Tobler K, Zhao Y, Weissman A, Majumdar A, Leong M and Shoham Z. **Egg Collection and Embryo Transfer Techniques.** *IVF-Worldwide.* 2010. | [Article](#)
- The use of gonadotropins and biosimilars in ART treated cycles. *IVF-Worldwide.* 2015. | [Article](#)
- Wiessman A, Yaron Y and Fishel S. **Preimplantation genetic screening (PGS): what is my opinion.** *IVF-Worldwide.* 2016. | [Article](#)
- Vaisbuch E, Leong M and Shoham Z. **Progesterone support in IVF: is evidence-based medicine translated to clinical practice? A worldwide web-based survey.** *Reprod Biomed Online.* 2012; **25**:139-45. | [Article](#) | [PubMed](#)

Citation:

Shoham G, Leong M, Weissman A and Yaron Y.
Can We Learn from the “Wisdom of the Crowd”?
Finding the Sample-Size Sweet Spot – an Analysis of
Internet-Based Crowdsourced Surveys of Fertility
Professionals. *J Med Stat Inform.* 2019; **7**:3.
<http://dx.doi.org/10.7243/2053-7662-7-3>