# Virtual Humans in Health-Related Interventions: A Meta-Analysis

**Tengteng Ma**
Dept. of Information and Decision Sciences
University of Illinois, Chicago
tma24@uic.edu

**Hasti Sharifi**
Department of Computer Science
University of Illinois, Chicago
hshari3@uic.edu

**Debaleena Chattopadhyay**
Department of Computer Science
University of Illinois, Chicago
debchatt@uic.edu

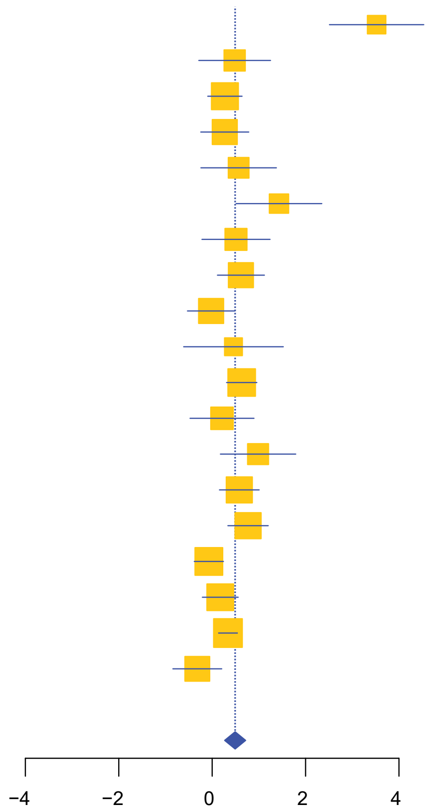


**Figure 1: Forest plot (*n* = 19) examining the effectiveness of virtual humans in improving health-related outcomes.**

**KEYWORDS**

Virtual humans; virtual characters; healthcare; avatars; meta-analysis; digital interlocutors.

## ABSTRACT

Virtual humans are computer-generated characters designed to simulate key properties of human face-to-face conversation—verbal and nonverbal. Their human-like physical appearance and nonverbal behavior set them apart from chatbot-type embodied conversational agents, and has recently received significant interest as a potential tool for health-related interventions. As healthcare providers deliberate whether to adopt this new technology, it is crucial to examine the empirical evidence about their effectiveness. We systematically evaluated evidence from controlled studies of interventions using virtual humans on their effectiveness in health-related outcomes. Nineteen studies were included from a total of 3354 unique records. Although study objectives varied greatly, most targeted psychological conditions, such as mood, anxiety, and autism spectrum disorders (ASD). Virtual humans demonstrated effectiveness in improving health-related outcomes, more strongly when targeting



*CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK.*



ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312853

**A Note on Vocabulary**

We use the term *Virtual Humans*. In the following list, we enumerate other commonly used terms in the literature to describe similar entities, but not *always* the same.

- Avatar/ Digital avatar
- Virtual characters/ Virtual humans
- Computer-modeled characters/ Computer-modeled entities
- Embodied conversational agent (ECA)/ Relational agent
- Virtual assistant/ Intelligent assistant
- Serious games/ Games for health

For example, an avatar often means a computer generated representation of the user, such as in video games, and controlled directly by the user's input. ECAs and relational agents simulate human turn-taking and use natural language generation algorithms but do not require a human-like physical appearance; they are mostly realized as chatbots. Virtual assistants like Siri or Alexa neither require a human-like physical appearance. Virtual characters may not be human but computer-animated cartoon characters. Serious games may use computer generated human-like characters, besides the user's avatar, but they are not required to simulate key properties of face-to-face conversation.

**Virtual Humans**: Computer-generated and computer-controlled characters with a human-like physical appearance and ability to simulate key properties of human face-to-face conversation, verbal (e.g., speech, voice modulation, turn-taking conventions) or nonverbal (e.g., gaze, emotions, head movement, metaphoric gestures) or both.

clinical conditions, such as ASD or pain management, than general wellness, such as weight loss. We discuss the emerging differences when designing for clinical interventions versus wellness.

## INTRODUCTION

Virtual characters and environments are rapidly becoming a popular tool for health-related assessments and interventions. As artificial intelligence (AI), computer graphics (CG), and ubiquitous computing technologies become increasingly powerful, it is worthwhile to take stock and examine how current applications are faring, and in turn, inform technology adoption and technology design decisions. But, since driving innovation is a key mission of the HCI community, sometimes empirical investigation of these innovative designs—particularly replication of such studies—takes a back seat. However, when striving for healthcare adoption, the norm is offering systematic empirical evidence of efficacy, which is ideally done using a meta-analysis.

A meta-analysis of the effectiveness of an emerging interdisciplinary technology is challenging, particularly when empirical studies are rarely replicated, technology designs are disparate and often not fit for comparison, and disciplinary priorities are divergent if not orthogonal. It is therefore not surprising that systematic reviews of virtual characters in healthcare is scant. Toward addressing that, we present a meta-analysis of a type of virtual characters—*virtual humans*—in health-related interventions (see the sidebar for a note on terminologies). While typical meta-analyses may include many studies, due to the issues mentioned above, this meta-analysis included 19 studies (Figure 2).

Virtual humans have long been used in video games, like Second Life, and more recently in serious games (or games for health) as interventions for healthy eating and nutrition. Serious games always offer a digital avatar, a virtual representation of the player–themselves or thereof (e.g., a pet), and may also include other virtual characters. But the efficacy of these interventions is primarily grounded in the game narrative, incentives, and manipulation of avatar characteristics to attenuate how much players identify with their avatars. This review does not focus on how the relationship between individuals and their avatars bring about health behavior changes. Instead, we examine the effectiveness of interactions between humans and virtual humans in health-related interventions. Besides effectiveness, there are other concerns in adoption of virtual humans in healthcare, such as ethical and legal issues, or perception and willingness of healthcare workers, which is being explored by healthcare researchers elsewhere.

Within health and wellness, virtual humans can help in senior care, e.g., providing options for home care monitoring and companionship, promote physical exercise, or offer alternatives to in-person psychotherapy. Although prior works have reviewed similar interventions, the focus had been on low-tech solutions that can deliver Internet-based mental health interventions for routine clinical practice [15]. With the rapid development of CG and AI, virtual humans in health-related applications have attracted a substantial amount of attention. However, comprehensive evidence about their

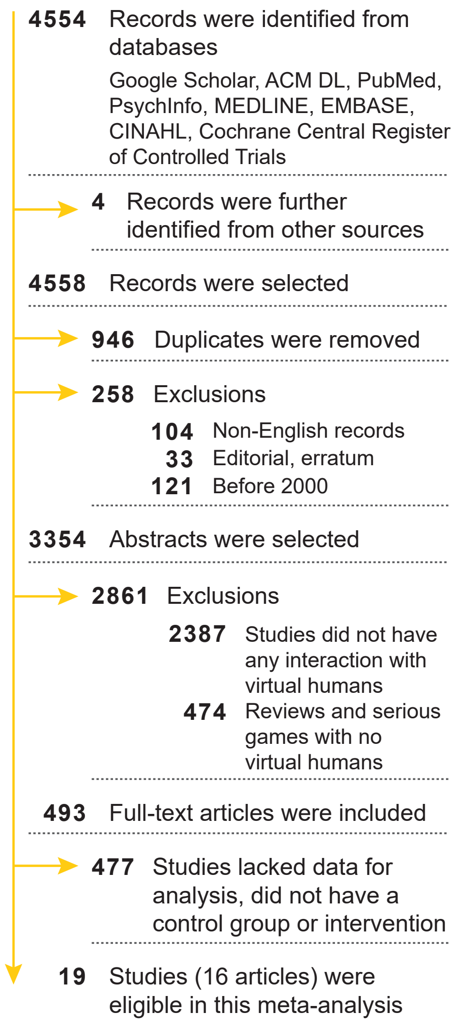


**Figure 2: Summary of literature search following PRISMA guidelines [13].**

efficacy remains lacking. We systematically review, synthesize, and quantify the currently available evidence on the efficacy of virtual humans in interventions across a range of health-related outcomes.

## METHODS

Besides systematic qualitative review, quantitative aspects play a key role in evidence synthesis. The statistical methodology for combining quantitative evidence from different studies is called meta-analysis, a staple in healthcare research. A meta-analysis allows synthesizing evidence from studies that have compared different interventions and measured these interventions with different outcomes. It combines the influential factors from all available works and quantifies the overall validity of the findings. Following standard guidelines for conducting meta-analyses, we devised a search strategy, established inclusion and exclusion criteria, extracted data, synthesized data, and performed a statistical analysis with eligible studies.

Focusing on virtual humans in health-related interventions, we identified two domains, (1) HCI and (2) healthcare, developed a set of keywords (Table 1) and identified a set of electronic databases to mine literature (Figure 2). The search was conducted between January and November 2018 using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [13]; all search queries were used separately in each identified database.

Studies were excluded if (1) not in English, (2) did not include a virtual human in a health-related intervention, (3) did not have a control group, or (4) did not report necessary statistics (*M* and *SD* of the outcome measure in control and experimental groups). Empirical studies were included where (1) the interaction between humans and virtual humans was a part of the intervention and (2) a health-related outcome was measured. The first and third author independently assessed the relevancy of search results and selected full-texts for further review. The first and second author independently abstracted the key study factors into a data extraction form and then came to a consensus on which studies met the inclusion criteria; the third author made the definitive decision when discrepancies would arise.

With our inclusion criteria, 19 studies from 16 articles were selected from a total of 3354 unique records, which were then synthesized using the PICOTS (population, intervention, comparison group, outcomes, time frame, setting) framework [16]. Because the included studies varied in terms of the population (from children to older adults), health-related outcome (e.g., improving social skills in ASD patients or help with substance abuse), and the intervention design (e.g., delivered via desktop, tablet, or virtual reality, Tables 2 & 3), a random effects model was used for the meta-analysis. Random effects model do not assume that the estimated effects come from a single homogeneous population, but the fact that true effect sizes vary from study to study. All statistical analyses were performed in R using the META package. Standardized mean differences (SMD) were used as the effect size to quantify the overall effectiveness of virtual humans in health-related interventions.

**Table 1: Search strategy and queries.**

| Domain | Search Terms |
|---|---|
| Avatars | (avatar OR embodied conversational agent OR computer-modeled entities OR digital characters OR digital interlocutors OR computer-modeled characters OR humanlike character OR anthropomorphic character OR cartoon character OR relational agent) |
| Narrative | (digital narrative OR digital health education OR digital plot OR health games OR games for health OR exergames OR healthy games OR active games OR active videogame OR active video game OR serious games OR health videogames OR story immersion) |
| Assistant | (digital assistant OR virtual assistant OR digital caregiver OR virtual caregiver OR virtual nurse OR digital nurse) |

**Some Characteristics of Selected Studies**

exclusively women as participants: 2 [1, 7]
exclusively children as participants: 2 [2, 19]
exclusively older adults as participants: 1 [9]
example of clinical conditions: overactive bladder (OAB) [1], depression [4], schizophrenia [5, 11], substance abuse [8], ASD [9], terminal care [19], Type 2 Diabetes [17], chronic stress [18].
example of wellness interventions: exercise [2], health education [3], lifestyle [6, 7, 12].

## RESULTS

All 19 studies selected for the meta-analysis were randomized clinical trials (RCTs), where a virtual human was used as the intervention; they were all published after 2010.

Study sample sizes ranged from 14 to 354. The total sample size was 1525, with 742 belonging to the intervention or experimental group. These studies can be broadly categorized into two types of design objectives: (1) for clinical interventions [1, 4, 5, 8–11, 14, 17–19] or (2) for general wellness, for instance, lifestyle changes [2, 6, 7] or informal health education [3, 12].

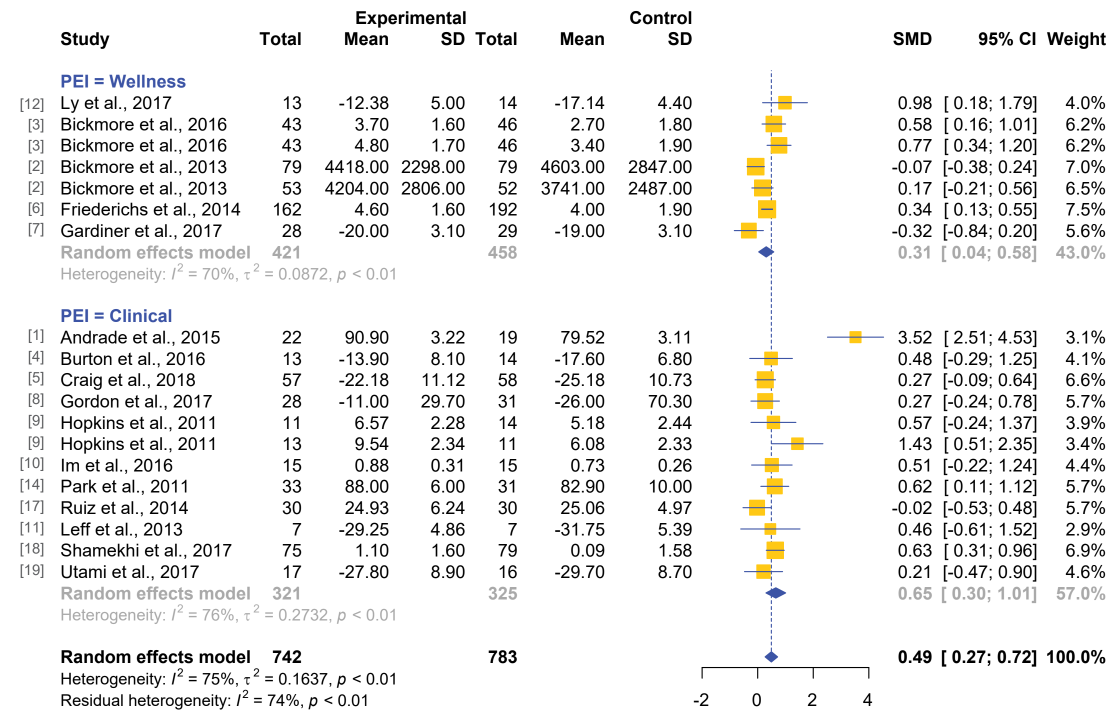


**Figure 3: A meta-analysis suggested a significant positive effect of health-related interventions using virtual humans compared with no virtual humans, *SMD* = 0.49, *p* < 0.01, 95% CI [0.27, 0.72]. A subgroup analysis revealed a larger effect for clinical interventions, *SMD* = 0.65, *p* < 0.01, 95% CI [0.30, 1.01], than interventions for general wellness, *SMD* = 0.31, *p* < 0.01, 95% CI [0.04, 0.58].**

### Overall Effectiveness

A meta-analysis suggested a significant positive effect of health-related interventions using virtual humans compared with no virtual humans, $SMD = 0.49$, $p < 0.01$, 95% CI [0.27, 0.72]. A subgroup analysis revealed a larger effect for clinical interventions, $SMD = 0.65$, $p < 0.01$, 95% CI [0.30, 1.01], than interventions for general wellness, $SMD = 0.31$, $p < 0.01$, 95% CI [0.04, 0.58].

### Interpreting the Meta-Analysis

- *SMD* or standardized mean difference is used as a summary statistic in meta-analysis when all the research studies assess the same outcome, but measure it in different ways.
- *SMD* is often used interchangeably with *effect size* and interpreted similar to Cohen's $d$: small, $SMD = 0.2$; medium, $SMD = 0.5$; and large, $SMD = 0.8$.
- *SMD* is a point estimate of the effect of an intervention. Calculating 95% confidence intervals (CIs) for the *SMD* facilitates comparing the effects of different interventions.
- The meta-analysis presented in this paper suggests that comparing with traditional interventions, interventions using virtual humans have more efficacy—with a medium effect size for clinical conditions ($SMD = .49$) and a small effect size for wellness ($SMD = 0.31$).

## DISCUSSION

Using a systematic review and meta-analysis, we showed empirically an effectiveness of interventions using virtual humans in health-related outcomes, thereby synthesizing evidence from currently available studies (see the sidebar for effect sizes). Our analysis has limitations, particularly with the wide variety in population, intervention, and outcome types (Tables 2 & 3). Furthermore, another obvious limitation is the scarcity of appropriate studies that use virtual humans as an intervention on health-related outcomes. Nevertheless, our work is a first step toward critically examining the evidence of effectiveness of virtual human interventions in healthcare.

It is encouraging that a subgroup analysis found a stronger positive effect of virtual human interventions in treating clinical conditions in comparison with general wellness (Figures 1 & 3). Although most designs were preceded with one or more requirements gathering phases, using focus groups, surveys, or interviews, it appears that virtual human interventions for specific clinical conditions were more tailored to known aliments and thereby more effective. Further studies are needed to unpack the design dimensions of virtual human interventions and factors that lead to intervention efficacy. For example, when treating clinical conditions, often the design decisions were heavily authored by medical specialists. It is notable that a majority of studies treating clinical conditions focuses on mental health. Given our positive results, virtual health interventions might catalyze a new era of virtual therapy, affordable and accessible to a wider range of individuals—who otherwise cannot get mental health treatments.

Future work will include analyzing the risk of bias and study quality, unpacking design decisions that led to intervention effectiveness, and reviewing virtual humans in health-related assessments. This line of inquiry can inform future design decisions for virtual health applications in healthcare and also provide evidence for healthcare adoption.

## REFERENCES


[1] Allen D. Andrade, Ramanakumar Anam, Chandana Karanam, Pamela Downey, and Jorge G. Ruiz. 2015. An overactive bladder online self-management program with embedded avatars: A randomized controlled trial of efficacy. *Urology* 85, 3 (2015), 561–567.

[2] Timothy W. Bickmore, Rebecca A. Silliman, Kerrie Nelson, Debbie M. Cheng, Michael Winter, Lori Henault, and Michael K. Paasche-Orlow. 2013. A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society* 61, 10 (2013), 1676–1683.

[3] Timothy W. Bickmore, Dina Utami, Robin Matsuyama, and Michael K. Paasche-Orlow. 2016. Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment. *Journal of medical Internet research* 18, 1 (2016), e1.

[4] Christopher Burton, Aurora Szentagotai Tatar, Brian McKinstry, Colin Matheson, Silviu Matu, Ramona Moldovan, Michele Macnab, Elaine Farrow, Daniel David, Claudia Pagliari, Antoni Serrano Blanco, and Maria Wolters. 2016. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. *Journal of Telemedicine and Telecare* 22, 6 (2016), 348–355.

**Table 2: Technology Characteristics**

| Delivery Modality | Studies | # |
|---|---|---|
| desktop GUI | [4–6, 9, 11, 17] | 6 |
| touch-enabled tablets | [1, 2, 18, 19] | 4 |
| web-based | [3, 7] | 2 |
| smartphone | [12] | 1 |
| virtual reality | [10, 14] | 2 |
| 3D desktop environment | [8] | 1 |

**Table 3: Nonverbal Behavior found in Selected Studies (9 out of 19 studies)**

| Nonverbal Cues | Studies | # |
|---|---|---|
| hand gestures | [2, 3, 18, 19] | 4 |
| gaze | [2, 3] | 2 |
| facial expressions | [3, 4, 9, 18, 19] | 5 |
| eye movements | [6] | 1 |
| head movements | [6, 19] | 2 |
| body movements | [8, 10] | 2 |
| posture | [19] | 1 |

**Study Limitations**

- Fewer studies for wellness interventions than clinical conditions met the inclusion criteria, potentially biasing the effect size.
- Between-study heterogeneity was substantial, $I^2 = 75\%$, possibly because of influencer cases like [1].


[5] Tom KJ Craig, Mar Rus-Calafell, Thomas Ward, Julian P. Leff, Mark Huckvale, Elizabeth Howarth, Richard Emsley, and Philippa A. Garety. 2018. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. *The Lancet Psychiatry* 5, 1 (2018), 31–40.

[6] Stijn Friederichs, Catherine Bolman, Anke Oenema, Janneke Guyaux, and Lilian Lechner. 2014. Motivational interviewing in a web-based physical activity intervention with an avatar: Randomized controlled trial. *Journal of medical Internet research* 16, 2 (2014).

[7] Paula M. Gardiner, Kelly D. McCue, Lily M. Negash, Teresa Cheng, Laura F. White, Leanne Yinusa-Nyahkoon, Brian W. Jack, and Timothy W. Bickmore. 2017. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling* 100, 9 (2017), 1720–1729.

[8] Michael S. Gordon, Steven B. Carswell, Mary Schadegg, Kayla Mangen, Kelly Merkel, Susan Tangires, and Frank J. Vocci. 2017. Avatar-assisted therapy: a proof-of-concept pilot study of a novel technology-based intervention to treat substance use disorders. *American Journal of Drug and Alcohol Abuse* 43, 5 (2017), 518–524.

[9] Ingrid Maria Hopkins, Michael W. Gower, Trista A. Perez, Dana S. Smith, Franklin R. Amthor, F. Casey Wimsatt, and Fred J. Biasini. 2011. Avatar assistant: Improving social skills in students with an asd through a computer-based intervention. *Journal of Autism and Developmental Disorders* 41, 11 (2011), 1543–1555.

[10] Hyungjun Im, Jeunghun Ku, Hyun Jung Kim, and Youn Joo Kang. 2016. Virtual reality-guided motor imagery increases corticomotor excitability in healthy volunteers and stroke patients. *Annals of Rehabilitation Medicine* 40, 3 (2016), 420–431.

[11] Julian Leff, Geoffrey Williams, Mark A. Huckvale, Maurice Arbuthnot, and Alex P. Leff. 2013. Computer-assisted therapy for medication-resistant auditory hallucinations: Proof-of-concept study. *British Journal of Psychiatry* 202, 6 (2013), 428–433.

[12] Kien Hoa Ly, Ann Marie Ly, and Gerhard Andersson. 2017. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions* 10, August (2017), 39–46.

[13] David Moher, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4 (2009), 264–269.

[14] Kyung Min Park, Jeonghun Ku, Soo Hee Choi, Hee Jeong Jang, Ji Yeon Park, Sun I. Kim, and Jae Jin Kim. 2011. A virtual reality application in role-plays of social skills training for schizophrenia: A randomized, controlled trial. *Psychiatry Research* 189, 2 (2011), 166–172.

[15] Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. 2017. Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research* 19, 5 (2017).

[16] John J Riva, Keshena MP Malik, Stephen J Burnie, Andrea R Endicott, and Jason W Busse. 2012. What is your research question? An introduction to the PICOT format for clinicians. *The Journal of the Canadian Chiropractic Association* 56, 3 (2012), 167.

[17] Jorge G. Ruiz, Allen D. Andrade, Ramanakumar Anam, Miriam Lisigurski, Chandana Karanam, and Joseph Sharit. 2014. Computer-Based Programmed Instruction Did Not Improve the Knowledge Retention of Medication Instructions of Individuals With Type 2 Diabetes Mellitus. *The Diabetes Educator* 40, 1 (2014), 77–88.

[18] Ameneh Shamekhi, Timothy Bickmore, Anna Lestoquoy, and Paula Gardiner. 2017. Augmenting group medical visits with conversational agents for stress management behavior change. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10171 LNCS (2017), 55–67.

[19] Dina Utami, Timothy Bickmore, Asimina Nikolopoulou, and Michael Paasche-Orlow. 2017. Talk about death: End of life planning with a virtual agent. In *International Conference on Intelligent Virtual Agents*. Springer, 441–450.